

Perils in free market genomics

Scientific and medical enthusiasm for the potential contributions of germline gene therapy must not obscure the need for detailed debate about its potential consequences — or for careful monitoring and sensitive regulation.

One of the most successful political initiatives of recent years, espoused by both right and left, has been the drive for deregulation. Rolling back the power of the state to let individual initiative and imagination flourish is no longer the exclusive rallying cry of conservatives, but has won a growing number of converts across the full political spectrum. But deregulation also has a price, as those who have experienced lengthy delays in US airport waiting lounges or the confusion over bus services in Britain are well aware.

So it has been with genetics. It was perhaps inevitable that scientific and industrial enthusiasm for the promises of genetic engineering should have also generated widespread concern — often exaggerated — about the potential dangers of their careless application, and that this should have been expressed through regulations covering what could be done, and under what conditions. Equally inevitable has been the resentment among researchers who feel unnecessarily restricted in what they are permitted to do as a result. But prudence is a virtue abandoned at peril.

As a meeting last week at the University of California, Los Angeles, illustrates, such issues are already raising their head over what many consider to be the ultimate target of such techniques: the ability to modify the human germ line at will (see page 317). It would be foolish to ignore the vast medical possibilities that this ability is already opening up, with its unprecedented potential for reducing human suffering; whatever the views of critics, public opinion polls already indicate the eagerness with which the promise of germline gene therapy is being greeted. It would be equally foolish to pretend that draconian regulations aimed at holding back the development of the relevant knowledge and skills are either desirable or likely to be effective.

But both factors make sensitive and comprehensive regulation

more, not less, necessary. The speakers at last week's meeting have performed a valuable service in opening up an area that is too often considered taboo. But, whereas some spoke of the dangers that research and its applications might be slowed down by public concerns, others were more cautious. Some warned, for example, of the unforeseen hazards that might emerge from tampering with the human genome, and that are likely to be avoided only by careful monitoring. Others spoke of the difficult task of establishing a boundary between acceptable and unacceptable applications. And serious concerns about the purely moral dimensions of deliberate intervention in the human germ line intended primarily to enhance culturally desirable characteristics cannot be dismissed lightly.

There is no reason to fear germline gene therapy as such. Those who argue that genetic enhancement is a process that has been carried out unconsciously for millennia have a point. But nor is there a reason to fear society's involvement in its careful monitoring and regulation. The medical and social revolution that the techniques are likely to bring are of such sweeping significance that it would be both foolhardy and dangerous to leave the development of the field to the whim of individuals. We share a collective responsibility to ensure that such developments take place in an acceptable manner.

Our first task should be to take a long, hard look at which is likely to be involved — both scientifically and ethically. Other important priorities include gathering extensive data on the long-term consequences of current experiments with somatic gene therapy, testing germline techniques thoroughly on other primates, and eventually assessing the potentially devastating consequences of failed experiments in humans. Only then will we be able to make the difficult choices about the best direction to take. □

All hands on deck?

Effective maritime research requires collaboration, but the case for a European agency has still to be made.

When the oil tanker *Amoco Cadiz* ran aground off the coast of France exactly 20 years ago, it awakened French public opinion on maritime issues in a way that found political expression in the creation of a 'ministry of the seas' after the presidential victory of François Mitterrand in 1981. France and Portugal are now leading calls for Europe as a whole to give greater political prominence to managing its marine environments and resources. They want to create a European Maritime Agency that would in particular forge closer links between maritime science and policy making in both government and industry.

The world beneath the waves certainly suffers from lack of political visibility. Moreover, the various actors tend to have a blinkered view of the issues, being preoccupied with those that affect them most directly in the short term. They often miss the bigger — and more complex — picture constituted by the vast sprawl of maritime research and related activities, as shown by meeting on maritime issues held last week in Paris, attended by more than 100 members of different European parliaments (see page 323).

Anything that might help to remedy this situation is welcome.

Much still needs to be learnt about the oceans, while many problems and opportunities — from the decommissioning of oil rigs to the promises of marine biotechnology — could benefit from a more imaginative approach to marine science. A suggestion that the proposed agency be steered by ministers from European countries would at least have the benefit of focusing political and public attention. But the main challenge facing those keen to promote such an agency is to identify where a 'European' approach could clearly advance maritime issues — not easy, given that many are either global or regional — without creating either another talking shop or a top-heavy bureaucracy.

Similarly, while the Europeanization of research resources should improve cost-effectiveness, this might well be achieved more simply through multilateral agreements. At present, the fog surrounding such questions suggests that the proposed agency could turn out to be a solution looking for a problem. But France and Portugal, which are championing the idea, deserve to be encouraged. Their intention to scrutinize maritime management should in itself prove a useful — and long overdue — exercise. □



Germline gene therapy 'must be spared excessive regulation'

[LOS ANGELES] Germline gene therapy is likely to become a reality within 20 years and should be welcomed, a high-level panel of scientists and other experts said at a meeting in the United States last week. They warned, however, that the procedure could suffer from excessive regulation at either national or international levels.

Any international attempt to regulate germline engineering would be "a complete disaster", said James Watson, the co-discoverer of the structure of DNA, who is the president of Cold Spring Harbor Laboratory on Long Island, New York.

Scientists should proceed unhindered towards germline engineering, he told a symposium called 'Engineering the Human Germ Line', organized by the University of California, Los Angeles. He added: "If there is a terrible misuse and people are dying, then we can pass regulation."

Most of the ten experts on the panel stressed that attempts at germline gene therapy must be preceded by extensive work in animals and human cell lines to develop techniques that would be safe and effective in human embryos. Possible dangers include harmful and unpredictable interactions between inserted or modified genes and others in the recipient genome, causing, for instance, cancer.

But the panellists almost unanimously argued that, once these concerns have been



Watson: research should go ahead.

Leroy Hood, chair of molecular biotechnology at the University of Washington, said: "We are using exactly the same kinds of technologies that evolution [does]." John Fletcher, a bioethicist at the University of Virginia, said that references to the germ line as a Rubicon not to be crossed, and as being "sanctified", had been virtually enshrined in public policy. "I think [this symposium] tended to dispute that premise."

In contrast to Europe (see panel), whose governments have already indicated that they will take a firm stand on the new technology, the United States has no law that would prohibit germline manipulation for whatever purpose, provided experiments passed safety and efficacy muster with the Food and Drug Administration. Privately funded research towards germline gene therapy using human embryos is also legal in the United States.

But the Recombinant DNA Advisory Committee (RAC) of the National Institutes of Health — which provides a public forum

addressed, the potential for curing human disease presented by the incipient technology is so great that it should be implemented — regardless of concern that its use might lead to an ethical morass, and perhaps even to practices such as eugenics.

for discussion of the ethical issues involved in gene therapy — has so far refused to consider germline gene therapy proposals. This may change. The RAC is updating its guidelines, which were written in 1990.

"This will be an opportunity for the committee to revisit" the germline gene therapy issue, says the RAC's director, Claudia Mickelson, who is the biosafety officer at the Massachusetts Institute of Technology. Mickelson says the RAC needs to make explicit the conditions under which it would consider such proposals, or to explain why it refuses to do so. She believes that the ethical issues involved need serious examination before such work proceeds.

A lone voice on the panel called for a sharp line to be drawn between germline therapy for enhancement and to fight disease. This call came from French Anderson, a professor of biochemistry and paediatrics at the University of Southern California School of Medicine, who pioneered human somatic gene therapy in 1991.

Anderson argued that, because the possible harmful effects of manipulating the germ line are unknown, researchers have a duty "to use this powerful technology [only] for the treatment of disease and not for any other purpose". He proposed that before germline therapy to fight disease proceeds in humans, long-term experience with somatic gene therapy in hundreds of patients must be accumulated over at least another decade; reliable, reproducible and safe procedures must be demonstrated in primates; and social awareness and approval must be gained.

Presenters at the symposium made it clear that the technology to conduct germline manipulations is rapidly approaching. For instance, human artificial chromosomes, that ultimately could carry hundreds of genes, are expected soon to be in use in somatic-cell gene therapy. And their use is expected to be made markedly easier by the advent of DNA chips.

The panellists said that germline therapy should be technically easier than somatic gene therapy, which in a decade of attempts has produced poor results. Because germline therapy aims at making changes in a single cell, the zygote, the procedure is "actually much simpler" said Mario Capecchi, a professor of biology and genetics at the University of Utah, who produced "knockout" mice by inactivating single genes.

But the panel's scientists conceded that the time to germline therapy is at least a decade off. "It's going to take a long time to work out the bugs," said Capecchi. **Meredith Wadman**

European states outlaw permanent changes

[LOS ANGELES] Gene therapy in humans has so far been attempted only in somatic cells, where genetic changes introduced in an individual die with that person. By contrast, germline gene therapy (see above) — which would be carried out on newly fertilized human zygotes — would introduce changes not only in every cell of the infant born of such manipulations but also in the genes passed to that baby's progeny. (Some of the scientists at the symposium said, however, that simple methods could keep changes in the recipient from becoming permanent.)

The potential for eliminating disease with such methods is tantalizing. For example, single-base changes in the DNA sequence cause devastating ailments such as Tay-Sachs disease. Such relatively simple targets would be prime candidates for early attempts at germ-line therapy. Complex diseases involving many genes would require much more research and are likely to be targeted farther into the future.

But critics are concerned about the potential for abuse of the technology to 'enhance' healthy individuals, by, for instance, manipulating intelligence, emotional stability, longevity

or physical appearance. They present the spectre of a booming market in 'designer' babies, made to the specifications of parents hungry for success and superiority in their offspring.

Such concern, in part, informed a bioethics convention produced by the Council of Europe last year that has since gathered signatures from 22 European states. This convention says that genetic manipulation may be undertaken for purposes of prevention, diagnosis or therapy — but only if does not aim to introduce a permanent modification in the genome. **M. W.**

Review could speed up German grants

[MUNICH] Germany's main funding body for university research, the Deutsche Forschungsgemeinschaft (DFG), is undergoing its first external evaluation. The result could be wide-ranging reforms — including perhaps the speeding up of the peer-review process — next year, the fiftieth anniversary of the DFG's re-establishment after the Second World War.

The idea of an extensive external evaluation was proposed by the DFG two years ago. A group of ten international experts was appointed by the Bund-Länder-Kommission, which coordinates federal and state governments' research policy. The group is headed by Richard Brooke, chief executive of the UK Engineering and Physical Sciences Research Council, and held its first discussions in Bonn last month.

The evaluation will focus particularly on the DFG's grants system. Ernst-Ludwig Winnacker, the new DFG president, is concerned that many grants last for too short a period, and that the reviewing procedure takes too long — and is out of line with European norms. In contrast to the United Kingdom and the United States, where major grant-giving bodies appoint their own referees, the scientific community elects the DFG's 500 or

so referees every four years. Winnacker asks whether "democracy [is] the best way forward to scientific excellence".

Most German researchers appear reluctant to abandon the democratic culture of the review system. A survey last year of 1,200 recent DFG grant applicants — plus a further 200 academics who had not applied for grants recently — found 80 per cent in favour of the election of referees.

But a substantial criticism raised by the survey was the length of time it takes the DFG to decide about grant applications — an average of about six-and-a-half months. Many researchers complain that this causes unnecessary uncertainty.

Working with the evaluation committee, the DFG is investigating ways of halving this time. Its somewhat random method for choosing referees for particular applications is seen as a significant factor in the delays, as those selected in this way are frequently unavailable. The DFG wants to guarantee the availability of a certain referee in future, using a computerized information system.

The agency also wants the evaluation committee to consider its at present modest programme for young scientists. This is a subject close to the heart of Winnacker, who

believes that young scientists need to be given responsibility for running their own research groups much earlier than at present, to encourage them to remain in Germany (see *Nature* 388, 507; 1997).

Last year, the DFG created a programme for young life scientists, allowing them to apply not only for research grants for a five-year period — instead of the normal two or three years — but also for their own salaries, which it does not normally allow. But only three such grants are available each year.

DFG officials are hoping that a recommendation from the evaluation group could help to expand the programme. Budgetary pressure has made it difficult for the organization to agree internally on expansion because even now only a low proportion of approved applications in other programmes can be funded.

But the DFG remains cautious about the effectiveness of the evaluation committee's report, due next year. "You can never predict if, and when, politicians are willing to match action to words," says Christoph Schneider, director for scientific and international affairs. Further inspections by and meetings of the evaluation group are planned for July and December.

Quirin Schiermeier

Russian church and scientists lay revolutionary quarrels to rest

[MOSCOW] Prominent Russian scientists and religious leaders have met to proclaim their common interests for the first time since before the 1917 Revolution. Unlike 1917, when religion was branded the "opium of the people", the scientists declared little conflict with religious thought, while their religious counterparts expressed concern about the lack of funding for science.

Both were attending a meeting called 'Faith and Knowledge: Science and Technology at the Frontier of Two Centuries', in Moscow last week. The Worldwide Russian National Council meeting was organized under the auspices of the Russian Orthodox Church and the Russian Academy of Sciences — a combination that would have been unthinkable only ten years ago.

The Russian Orthodox Church's patriarch for Moscow and the whole of Russia, Alexsij the Second, said: "Russia is the great scientific state, but the present economic crisis has damaged its scientific and technological potential, which means that in the coming century the country will face difficult obstacles."

He added: "Russia's fate is now in the hands of scientific intelligentsia, and whether or not these people are ready to mobilize their abilities and strengths to



serve Russia could not but be our church's concern."

The meeting represented the first time for at least 80 years that scientists and clergymen have met in Russia. During the Soviet era, religion was officially discouraged, and one of the major tasks of the Academy of Sciences was to promote anti-religious propaganda.

The situation today is very different. "Science is not in conflict with religion, and religion is also based on rationality, it's a kind of rationality," said Yuri Osipov, the president of the Russian Academy of Sciences. "A process of convergence is now

taking place between science and religion; they interact in building the human-oriented values of our culture."

Other speakers supported Osipov by pointing out that more than 40 per cent of scientists now openly call themselves believers — formerly all were considered to be atheists — whereas many clergymen hold scientific degrees from universities, another unprecedented step. Some scientists even compared the act of scientific discovery to a religious experience.

Vladimir Fortov, vice-president of the Academy of Sciences and until this week minister of science and technologies, said that science and religion have much in common; only their methods of understanding the natural world are different. The Big Bang theory, for example, is close to the theological view on the origin of the Universe, he said.

But the Russian Orthodox Church is not entirely enthusiastic about modern science. It strongly opposes any research on cloning not only humans but animals in general, and is suspicious about many other fields of science. "Nuclear weapons, artificial intelligence, information systems and genetic engineering, all contain a 'phantom' of enormous danger," said the Metropolitan Bishop Kirill.

Carl Levitin

MAFF in a stew over food research plans

[LONDON] Food safety researchers in Britain are locked into a bitter controversy over plans to restructure several laboratories. The main trigger has been a proposal by the Ministry of Agriculture, Fisheries and Food (MAFF) to close its main food safety laboratory in Norwich, and to relocate staff to the newly built Central Science Laboratory (CSL) 200 miles north in York.

These proposed changes coincide with uncertainty about the future of two other food laboratories — in Reading and in Norwich — that are part of the Institute of Food Research, run by the Biotechnology and Biological Sciences Research Council (BBSRC). An internal review of their operation suggested merging the two laboratories on one site.

MAFF's plans for the Norwich laboratory are being vigorously opposed by its scientists and many independent food researchers, as well as by consumer groups. Representatives of all three say they have difficulty in understanding the ministry's haste in ordering the changes, given that responsibility for food safety will soon fall into the hands of an independent Food Standards Agency.

MAFF officials say the move is needed to fill laboratory space at York — half of which lies vacant because of cuts to the ministry's research budget. The empty space has to be accounted for as all government departments now have to pay rent on property they own. "We can't afford to have empty buildings," say the officials.

The move is also being officially justified on the grounds that the average cost per scientist at York will be less than at Norwich. The York laboratory was built in mid-1996. The aim was to collect under one roof MAFF's 14 smaller food research laboratories. The new laboratory is engaged primarily in producer-oriented research, such as research into pesticides.



The front line: a succession of 'food scares' has made food safety research a contentious issue.

One option being considered by ministers is to close the BBSRC Reading site and move its work to the vacant MAFF building in Norwich — if the MAFF Norwich operation closes.

The local member of parliament, Charles Clarke (Labour, Norwich South), is supporting MAFF scientists in their bid to keep the laboratory in his constituency. He has tabled parliamentary questions to the food minister, Jeff Rooker, asking for a detailed breakdown of the relative costs of Norwich and York. "I haven't received any answers as yet," he says.

Clarke says he also intends to raise questions about the fate of the 'missing millions' — a reference to the mismatch between the York laboratory's £130 million (US\$217 million) building costs and its present value of £87 million.

Most researchers at Norwich are bitterly opposed to the move. But they fear that a decision may already have been made, in which case many — including senior staff — will probably resign. An offer of a management buyout is understood to have been rejected by civil servants. And the laboratory's director, John Gilbert, a known critic

of the plans, has been moved to a new MAFF post in London. He was out of the country and so unavailable for comment when this article went to press.

Earlier this month, Richard Packer, MAFF's permanent secretary and the department's highest-ranking civil servant, met staff at Norwich to explain the rationale behind the relocation plans. "Most at that meeting made clear their total opposition to the plans," says one scientist who declines to be named.

"These plans make no sense," says another scientist. "Norwich has made a name for itself as a centre for research into the safety of chemicals in food. Destroying such a facility should not be the answer to the financial problems of York."

The move is also being questioned by food safety experts and consumer groups, which argue that the ministry should be scaling down its direct involvement in food safety research, given that it will no longer be responsible for food safety following the government's decision to create an independent standards agency (see *Nature* 389, 109; 1997).

MAFF is meant to be withdrawing from the process of setting food safety research priorities even before the food standards agency is set up (see *Nature* 391, 313; 1998). "Any decision on the fate of the Norwich labs ought to be taken by the food agency, rather than MAFF," says David McWeeny, a former director of the Norwich laboratory.

Erik Millstone, a food policy researcher at the Science Policy Research Unit at the University of Sussex, agrees. He acknowledges that the agency will need a dedicated laboratory that can provide "firefighting" services in response to potential food scares. But the agency will not want to be permanently tied to one research facility as it will want to commission research from laboratories independent of government. "The agency needs the freedom to commission research whenever and wherever it likes."

But a MAFF official points out that the government only wants the agency to control the ministry's £25 million research budget — not the actual research laboratories. "The agency will have the freedom to decide on research [priorities]. But it cannot make decisions about a facility it does not own."

Consumer groups are concerned about the implications of Norwich's consumer-oriented research being housed alongside existing producer-oriented work at the CSL. Julie Sheppard, senior public affairs officer at the Consumers' Association, says that the move would be in conflict with the government's stated promise to place food safety research away from the influence of the food industry.

Ehsan Masood

Novartis goes public with fraud dismissal

[BASEL] The Swiss pharmaceutical industry appears to be taking a lead from public research institutions in Germany in making information about potentially embarrassing situations involving scientific fraud more readily available to the public.

Last week, Novartis reported that a scientist had been dismissed from its research department in Basel. In a recent routine internal investigation, the company found the researcher had continuously manipulated preclinical results in a cancer research project.

The research project is being conducted with California-based Isis Pharmaceuticals, and includes testing anticancer compounds

on xenografts of human tumours in animals. A spokesman says that this is the first such case to have emerged since Novartis was created two years ago by the merger of two large Swiss drug companies, Ciba and Sandoz.

The company is planning further investigations to determine what it describes as the "total damage" caused by the scientist. But it says it is not considering any additional routine control of researchers. Novartis and Isis will continue the joint project. "We have sufficient data and tests which are unaffected by the incident," says the spokesman.

Quirin Schiermeier

India awards science portfolio to advocate of nuclear arms

[NEW DELHI] Murli Manohar Joshi, a physicist, has been appointed cabinet minister for education and science and technology in India's new coalition government led by the Bharatiya Janata Party (BJP). Joshi, 64, believes in modernizing India through indigenous rather than imported technologies, and has also demanded that India should develop nuclear weapons.

The BJP won the largest number of parliamentary seats in the recent elections. The party is not known as a supporter of large-scale or expensive scientific projects, but this is the first time that the science portfolio has been given to a minister of cabinet rank; in the past it was handled by a junior minister.

It is also the first time that education and science portfolios have been allocated to the same minister. Assisted by a minister of state, Joshi — former professor of physics at Allahabad University and a past president of BJP — will look after all scientific departments apart from space and atomic energy. These, as in the past, will remain the direct responsibility of the prime minister, Atal Behari Vajpayee.

Vajpayee's national agenda says that his government "will re-evaluate the nuclear policy and exercise the option to induct nuclear weapons". Whether this means India will start building nuclear weapons is unclear.

But Rajagopala Chidambaram, chairman of the Atomic Energy Commission (AEC) and a key figure behind India's first nuclear test in 1974, has said that the country "has the technical capability" to do so. A former chairman of the AEC, Raja Ramanna, was also quoted recently in the *Times of India* as saying that India should go nuclear.

Neighbouring Pakistan says the BJP's nuclear policy is extremely worrying. Tariq Altaf, a spokesman for Pakistan's foreign office, said last week that Pakistan would be reviewing its policy of "nuclear restraint". He called on the international community to "take serious note of India's intentions". Pakistan is widely believed to be able to make nuclear weapons. Both India and Pakistan have refused to sign the Comprehensive Test Ban Treaty.

Strategic analysts fear new economic and foreign policy pressures for India. But Vajpayee says the new government will not go back on India's commitment to the World Trade Organization (WTO), including proposed changes to its patent law. Although contractual obligations would be met, he says, the government "would do some hard bargaining at WTO to ensure that national interests are better served". **K S Jayaraman**

New Zealand's universities face 'privatization' bid

[SYDNEY] Radical proposals from the New Zealand government for turning universities into private bodies fully exposed to market forces, and with all academics designated as either teachers or researchers, have generated fierce criticism from the academic community.

The idea was floated in a green paper (discussion document) last year. The government argues that the changes are needed to improve the effectiveness with which universities are run, and to ensure that they meet social and economic needs. It used the same argument when it created Crown Research Institutes out of the former Department of Scientific and Industrial Research (see *Nature* 391, 426; 1998).

But the country's Association of University Staff (AUS) has obtained the backing of 16 international affiliated organizations in a resolution to "condemn" the government for its "efforts to convert universities into business entities".

The International Conference of University Teacher Organizations (ICUTO) will shortly place New Zealand on a 'grey list' to alert academics contemplating taking a job at, or cooperating with, a New Zealand university to what it describes as "the deterioration of working conditions, collegial governance and academic freedom".

The main focus of discontent is a proposal to transform the management of universities from self-governing councils, with strong academic representation, into companies ruled by boards appointed by government to ensure compliance with "national goals".

Bryan Gould, chair of the New Zealand Vice-Chancellors' Committee and vice-chancellor of the University of Waikato, says it is "inappropriate and virtually unparalleled in the advanced world" to treat universities as private-sector corporations. "Their purpose is not to make profits for shareholders but to serve wider and longer-term aims."

The vice-chancellors' committee says that "restrictive government controls would be seen within New Zealand and abroad as

undermining the autonomy of institutions and their role as critic and conscience of society". The green paper suggests that the role of critic should be "relaxed".

Jane Kelsey, AUS academic vice-president and professor of law at the University of Auckland, says that publicity through ICUTO could generate an international test case, as 60 per cent of New Zealand university staff are recruited from overseas.

The impact of the government's proposals on the quality of research and teaching figures high among concerns expressed in submissions from five major science and academic bodies. The most controversial notion is the option of splitting academics into teachers and researchers, each with distinct sources of funding.

To achieve this, officials are suggesting that funds covering research and research-based teaching — estimated by the vice-chancellors' committee to be about 30 per cent of 'block grants' — should be transferred to a competitive fund run by the government.

Critics of such a move include the Royal Society of New Zealand, which attacks the review's "regard [of] education and knowledge merely as tradeable commodities", and opposes the proposed financing of universities through student demand expressed through portable 'vouchers'. The society predicts that this would favour subjects leading to "personal gain and perceived status", to the detriment of science.

The government is already moving to allow some polytechnics to apply to become universities. The green paper suggests giving private and overseas institutions equal access to funding through student vouchers. Universities are worried that such a move could cut their funding.

George Petersen, president of the academy council of the Royal Society of New Zealand, points to the 1990 Education Act's clear definition of a university as including research interwoven with teaching. "The paper contradicts this by seeming to drag universities and polytechnics to a lowest common denominator," he says.

There has also been criticism from the government-financed and appointed Foundation for Research Science and Technology.

The government had said that, once comments on the green paper were assessed, a white paper (policy document) would be published later this year. However, a senior official indicates that this may not happen and universities are concerned that, if the government bypasses this stage, decisions on the reforms will emerge piecemeal, starting with the budget in late May. **Peter Pockley**



Gould: universities serve wider and longer-term aims than making profits for shareholders.

Japan's NASDA under fire on launch failure

[TOKYO] The strategy of Japan's National Space Development Agency (NASDA) needs to be overhauled after the failure of three space missions in nine months, says Sadakazu Tanigaki, director-general of the Science and Technology Agency.

The space agency came under fire at a meeting held by Japan's Space Activities Commission over NASDA's slow investigation of the recent failure of an H-II rocket launch. The launcher, carrying the world's largest telecommunications satellites, failed to reach geostationary orbit after an engine failure (see *Nature* 391, 832; 1998).

The failure came as a severe shock to NASDA, whose two previous missions, the Advanced Earth Observation Satellite and the Advanced Land Observation Satellite, also ended in failure (see *Nature* 388, 105; 1997). An investigation has revealed an abnormality in the combustion chamber of H-II's second-stage LE5A engine, manufactured by Mitsubishi Heavy Industries. But the precise cause of the failure is still unknown.

Toshihiko Fujita of NASDA, who is responsible for the investigation, suggests the construction chamber may not have been strong enough. But he says it is unlikely there was a flaw in its design, pointing out that NASDA's past 30 medium-sized and large rockets have all been launched successfully.

Part of the problem appears to lie in the failure of NASDA to communicate adequately with the contractor who built the engines. Although the agency has a scheme



In happier times: a Japanese H-II rocket blasts off successfully from Tanegashima Space Centre.

to send its engineers to contractor companies, only one engineer has gone so far.

Although the immediate concern is to find the cause of the engine failure, others fear that the prolonged investigation may deter Japan from entering the international commercial satellite business.

Rocket System, a commercial satellite-launching company collaborating with more than 70 space-related companies, has signed a contract with Hughes Space and Communication International (HSCI) and Space Systems/Loral to launch more than 20 satellites using a new H-II launch vehicle. The

total cost of the operation, planned to happen between 2000 and 2007, is thought to be about ¥200 billion (US\$1.5 billion).

The two US commercial satellite manufacturers have so far shown little concern about the failure of H-II. Robert Berry, president of Space Systems/Loral, says "failures are indispensable to space development".

Although all H-II rockets have previously been developed indigenously, the new H-IIA will, as part of NASDA cost-cutting, be built using imported materials and subsystems. This is expected to reduce the cost from ¥18 billion per launch to less than ¥8.5 billion. "Now that Japan has an established technology for rocket development, the next step is to minimize the cost to enter global competition," says Hiroshi Imamura, director of Rocket System.

Imamura says inspection can be automated, and the construction of the rockets, which normally takes three years, can be reduced to two. The ultimate goal, he says, is to mass-produce rockets to launch them more frequently — only one H-II is, at present, launched each year in Japan.

With just two-and-a-half years left until Japan joins the international commercial satellite business, Imamura suggests that mass production of rockets could help improve mechanical reliability. "Since we only build one or two rockets each year, it is difficult for us to predict reliability, and increased production could help us improve our technology," he says. **Asako Saegusa**

Set clear priorities to win surplus cash, Canada's scientists told

[OTTAWA] The budget may have ended 14 years of deficit cutting and freed Canada to consider new science initiatives (see *Nature* 392, 7; 1998). But if science is to profit, hard choices have to be made and clear priorities set, as there will be stiff competition for the C\$70 billion (US\$49 billion) surpluses expected over the next three years.

This theme emerged from a colloquium on federal government support for science organized by the University of Ottawa's Program of Research on International Management and Economy (PRIME).

Robert de Cotret, a former president of the government's Treasury Board and now a faculty member of the university, spoke of February's budget as a "watershed" that enabled the government to consider projects formerly unthinkable. De Cotret also said the present government had shown a strong commitment to science and technology.

But other speakers agreed with Howard Alper, the university's vice-rector of research, who said that "putting money on the table will accomplish nothing" for

science. Specific proposals were needed.

Alper's own proposals included:

- more collaboration between the three research councils that provide grants — what he called "tri-council partnerships";
- focusing research priorities on areas important to economic development, such as biopharmaceuticals, materials, information technology, and food and agriculture;
- making strategic investments in international science and technology;
- allowing cash contributions by foreign companies to university–industry partnerships, partly to promote job creation;
- developing a programme to repatriate established, mid-career Canadian scientists now working abroad.

Dan Lane, the university's vice-dean of administration, complained that a lack of direction in both the fisheries and oceans and environment departments had led to confusion about goals for science.

There was general praise for the budget's restoration of council funding to 1995 levels and the establishment of the Millennium

Scholarship Fund. But J. Stephan Dupré, president of the Canadian Institute for Advanced Research, called the councils' increases mere "damage control".

Dupré said a big question was whether provincial governments would provide enough operating support for universities, citing scientific manpower shortages which universities were unable to correct.

Further caution was added by Don McDiarmid, of the Canadian Association of Physicists, who warned delegates not to expect industry to invest in basic research because it is not in its own interest. And Gilles Paquet, professor of public policy and management at the university, said that he saw "no action" on science in the budget and no evidence that Ottawa was seriously "looking at anything except infrastructure".

Moreover, he said, Canadian scientists were not helpful. "We still have a scientific community that is to a great extent corrupted by [the view that government should simply] send the money and not ask any questions," Paquet said. **David Surgeon**

Ecologists row over lobbying strategy

[WASHINGTON] Two letter-writing campaigns by ecologists intended to influence congressional action on the Endangered Species Act have touched off a heated debate about the advice scientists should give lawmakers.

Several authors of a letter sent in January to Senate sponsors of a bill to reauthorize the act (see *Nature* 391, 829; 1998) have criticized a second letter signed by the presidents of the Society for Integrative and Comparative Biology (SICB) and eight other scientific organizations, which ran as a full-page advertisement in *The New York Times* last month.

The SICB letter was organized by its conservation chairman, Fraser Shilling of the University of California at Davis, and takes a harder line on species protection than most of the scientists who advocate reform of the act. It calls for a complete ban on killing any endangered animal or plant, including the 'incidental takes' permitted for landowners who file habitat conservation plans.

Even though many conservationists worry that such plans are too inflexible, and sometimes based on inadequate science, they have become the primary means of setting aside privately owned land for conservation.

Frances James, a biologist at Florida State University who is leading a review of habitat conservation plans by the American Institute of Biological Sciences, was so upset with the *Times* advertisement that she persuaded

half a dozen colleagues to sign a letter chastising the signatories of the Shilling letter. The latter include the presidents of the Ecological Society of America, the Botanical Society of America and the Entomological Society of America.

James wrote that their action was "causing consternation among many biologists who have been working with endangered-species issues". She described early drafts of Shilling's recommendations — which had been circulating among scientists for a year — as "counterproductive and naïve".

Challenging ten specific points, she wrote that the blanket opposition to 'incidental takes' was "not scientifically valid", as it would prohibit valid conservation measures, such as the prescribed burning of habitat and control of invasive species. "We were sorry to see your letter in *The New York Times*, because it means that scientists will probably have to disagree in public about endangered species protection," she wrote.

Joining James were Gary Meffe of the University of Florida, Stuart Pimm of the University of Tennessee, and Peter Brussard and Dennis Murphy of the University of Nevada at Reno, all of whom signed the more moderate January letter to Senate leaders.

Murphy agrees that "we don't want a split in the scientific community". But he is not shy about criticizing the SICB letter which he

says "reads like the Sierra Club [conservation lobby group] wrote it instead of a group of scientific societies". He says the James letter is "an attempt by those who are doing their homework on this issue to police themselves and their colleagues".

Shilling says the criticism has less to do with science than politics. He calls the dispute a "territorial battle" between scientists too quick to yield to property rights advocates in the reauthorization battle and those — like himself — who would first "figure out what the standards [for species recovery] are, and then go in with our armour girded". Murphy counters by saying it is "not logical to ask a Republican Congress to make the act more stringent by orders of magnitude".

Although no signatories of Shilling's letter have retracted their endorsement, the Ecological Society of America considered withdrawing just before publication, because of the 'incidental take' clause and other objections. Another society president who signed admits: "We may have been caught by not being experts in this area, and by just feeling in general that we wanted to support a good cause."

To some, the incident illustrates how organized advocacy still does not come naturally to many scientists. "Scientific societies weren't put together as lobbying organizations," says Murphy.

Tony Reichhardt

Telescope users win respite from mobile phone interference

[WASHINGTON] Radioastronomers using the giant Arecibo telescope in Puerto Rico have won a promise from the owners of the Iridium satellite network to limit radio interference from the orbiting constellation for part of each day.

The agreement, which took five years to negotiate, guarantees eight hours of 'quiet' time from 22:00 to 6:00 Eastern time, during which time the Iridium system will not interfere with astronomical observations at the key frequency of 1612 MHz.

Arecibo officials hail the agreement as a good compromise. But other radio-astronomers worry that the interference issue is only getting worse.

The frequency band used for Iridium's 'downlink' to portable telephones on the ground is very near the emission frequency of the hydroxyl molecule, one of the most common interstellar molecules (see *Nature* 380, 569; 1996). Arecibo, the world's largest radio dish, is particularly suited to studying hydroxyl emissions, and some astronomers lobbied for unrestricted access to that band.

But, says Paul Goldsmith, director of Cornell University's National Astronomy and Ionosphere Center, which operates



Quiet please: the Arecibo telescope will have eight hours 'peace' a day after agreement.

Arecibo, scientists got the best deal possible. "Some radioastronomers may have felt that they were entitled to 24 hours a day, but I'm happy that both sides could agree to eight," he said.

The Iridium system is scheduled to begin operation this autumn, providing telephone services to users around the world. Other proposed networks will not downlink at the same frequency, and so do not pose as big a threat to radioastronomy.

Organizations that rarely observe hydroxyl emissions have found it easier to solve the Iridium problem. The US National Radio Astronomy Observatory (NRAO) operates the Very Large Array in New Mexico and the Green Bank telescopes in West Virginia, and needs only four hours of quiet time a day. It signed an agreement with Motorola, Iridium's owner, several years ago.

European radioastronomers are working out their own agreement with Iridium, and it may be even more difficult to negotiate than the Arecibo pact. The Nançay telescope in France, for example, spends about half its time studying hydroxyl emissions, and may need more than an eight-hour quiet period.

Tomas Gergely, who handles spectrum management issues for the US National Science Foundation, says each observatory has its own needs, which makes it difficult to work out a blanket agreement between satellite operators and the astronomy community as a whole.

In general, though, scientists face an uphill battle as more satellite systems like Iridium come online. "We must be very vigilant so that radioastronomy as a discipline survives," says Gergely.

T. R.

Politicians support maritime agency plans

[PARIS] More than a hundred European politicians last week gave tentative support to a proposal from France and Portugal to create a European Maritime Agency, responsible for coordinating a cross-sector approach to research and exploitation of maritime resources.

A meeting organized in Paris by the parliamentary assembly of the Council of Europe broadly agreed that the economic and scientific importance of maritime issues justifies greater political attention and a more forward-looking approach by Europe as a whole. But there seemed little agreement on the precise goals of the proposed agency, and widespread concern that it should not create unnecessary bureaucracy.

The economic and political stakes are high, given that maritime activities include oil exploration, fisheries, shipping, coastal management, pollution control and climate change. "We underestimate the maritime component of Europe," said Paul Roncière, secretary general of the General Secretariat on the Sea, a committee attached to the office of the French prime minister, Lionel Jospin. "Europe needs a distinct maritime policy."

The idea of a European Maritime Agency has been around for some time (see *Nature* 373, 553; 1995). But it has gained new impetus from vigorous support by José Mariano Gago, Portugal's minister of science and technology. Gago told the meeting that its creation would be a "major political step", arguing that "there is no European policy or body charged with coordinating or exploiting [maritime research]".

Claude Allègre, France's science minister, supported the proposals, arguing that in terms of basic science the oceans were still poorly understood. The UK science minister John Battle agreed that the challenges of ocean technology are "as demanding as those of space", but did not express a formal position on the idea of a new agency. But the United Kingdom is said to be open to the idea, if a realistic role can be demonstrated.

Sources say Battle and Allègre last week invited Gago to write to other European ministers with a detailed proposal on the agency's goals and structure. Britain also pressed for this to be accompanied by analysis before the summer by a group of experts from several European countries. Feedback from the meeting indicates Spain, Italy, Iceland and Norway may support the idea.

Draft proposals suggest that the agency's mandate would be decided by regular ministerial conferences, as with the European Space Agency (ESA). But the parallels with ESA end there; there seems broad agreement that maritime issues are too heterogeneous to lend themselves to management by a centralized operational agency, and that any new

body should be a lightweight structure for coordinating scientific, industrial and political activity.

A draft report to the Council of Europe's committee on science and technology by Pedro Roseta, a Spanish MP and the committee's rapporteur, proposes that the agency should study major maritime challenges, "conceive and if necessary manage research programmes"; plan the construction of large European research facilities and present a single European voice in international research projects.

But many feel that maritime research is itself relatively well coordinated. Many bodies already exist, such as the Intergovernmental Oceanographic Commission of the United Nations Educational, Scientific and Cultural Organization, and the International Council for the Exploration of the Seas. At the same time, many speakers, and politicians in particular, saw a need for a forum to give policy-makers and industrialists a means of obtaining scientific assessments of key areas in decision-making, such as coastal management and decommissioning of oil rigs.

The coordination of marine research within Europe has been markedly improved, say many observers, by the creation in 1995 of the European Marine and Polar Science Boards (EMAPS), a non-governmental organization set up under the auspices of the European Science Foundation — which represents most marine science organizations (see *Nature* 377, 469; 1995).

The embryo of a European maritime research programme was also created with the launch of the European Union's ECU250 million (US\$271 million) Marine Science and Technology Programme in 1989. Marine sciences were also a priority in the original proposal for the EU's next five-year Framework research programme, but have now been split up within other themes, creating a need for cross-sector coordination.

Given such structures, many are sceptical of the need for a new agency. John Krebs, chief executive of Britain's Natural Environment Research

Council (NERC), says: "From NERC's view there is not a great deal of need for another European body [in terms of research]."

But Pierre Papon, former head of IFREMER, the French marine technology research agency, argues that research capacities in Europe are too "dispersed" and "thinking is too national". He believes improvements could come from pooling resources and more strategic thinking, and cites the lack of strong research in marine biotechnology. Krebs agrees that "Europe has not made a big push" in this area, compared with the United States and Japan.

EMAPS is expected to release a report soon calling for a European initiative in marine biodiversity research. Such grassroots calls need a political outlet, says Laurent d'Ozouville, scientific secretary of EMAPS.

Allègre pointed out that, whereas France, Britain and Germany have agreed to cooperate in the use of the 40 or so existing research vessels (see *Nature* 379, 576; 1996), there is a need to think about cooperating to build new vessels. The only examples so far are the Franco-Spanish *Thalassa*, and a smaller Franco-Italian ship, *L'Europe*.

Krebs agrees the national approach to research ships is bound to give way to thinking about a "European fleet of research vessels". But he says that might be possible simply through multilateral agreements.

Another potential role for the agency that emerged at the meeting was to give Europe a single voice in international research programmes. In broad terms, the advantage of an agency might be to provide a "single point of contact" for a cross-sector approach to maritime issues, says d'Ozouville.

Coastal management, for example, involves research bodies, terrestrial and marine industries, and political action at regional, national and European levels, but "establishing that dialogue has been extremely difficult". "It needs European coordination and a multidisciplinary approach to provide scientific tools for the users managing these zones," says d'Ozouville. **Declan Butler**



Floating an idea: joint research ships such as the Franco-Spanish *Thalassa* may be a way forward.

Report said to seek more research in French universities

[PARIS] Research should be given a much greater role in French university courses, according to leaked details of a report being prepared for the government on the reform of universities and *grandes écoles*. The main changes suggested, reported in the newspaper *Libération*, include an ambitious reorganization of the French university degree structure to match that of other European countries. There are no suggestions of changes in the elite system of *grandes écoles* themselves.

The current two-year diploma of general universities studies would be extended by one year, and where appropriate would include professional and research training. A higher degree awarded after a further two years, equivalent in duration to the degrees now awarded by the *grandes écoles*, would similarly include six months of research – with a PhD being awarded after eight years of study.

Science minister moves on... after 20 years in job

[LONDON] Indonesia's minister for research and technology has moved to a new job after more than two decades in the post. B. J. Habibie has become his country's vice-president; his replacement, Rahardi Ramelan, was announced last week.

Habibie was an unusually powerful research minister thanks to his close links with President Suharto, a position that helped him to secure large sums of money for a number of ambitious technology projects.

Habibie's most notable achievement was to build a civil aircraft industry from scratch. But his bid to develop an indigenous high-technology base faltered when he had to relocate much of the industry to the United States to take advantage of lower costs of production.

Jospin pledges key role for basic research

[PARIS] The French prime minister, Lionel Jospin, has promised to put basic research at the centre of his government's programme. He warned last week that, although he wants to "bring the research world and economic spheres more closely together", he considers it "dangerous" for scientific research to be driven only by short-term economic and social demands.

Jospin was speaking at a reception in Paris organized by the Ecole Normale Supérieure and the Centre National de la Recherche Scientifique to celebrate the award of the Nobel prize in physics last autumn to Claude

Cohen-Tannoudji. Jospin, who is due to announce a national research strategy next month, said research would remain a priority in the state budget.

Israel to boost numbers of high-tech graduates

[JERUSALEM] Israel's prime minister, Binyamin Netanyahu, has ordered a sharp increase in the numbers of graduates in high-technology disciplines. The move follows a meeting last week with industry leaders who urged him to double the number of professionals in high-technology industries and to double those industries' exports.

The government had planned to raise the annual output of such graduates from 2,000 to 4,000 by the year 2003. But Netanyahu said such an increase is insufficient and he has asked his officials to submit a revised plan next month. Israel already has 15,000 high-technology professionals, and its industry exports US\$7 billion worth of goods and services annually.

Stars rally to support biomedical agency

[WASHINGTON] Hollywood was again enlisted last week to advance the medical research cause, with actors Christopher Reeve and Mary Tyler Moore appearing at a rally on Capitol Hill. The actors, a dozen politicians and other supporters of the US National Institutes of Health (NIH) repeated a call to Congress to double the budget of the \$13.6 billion agency in the next five years.

Reeve, who was paralysed in a riding accident in 1995, said that the United States is at the point of being able to buy cures and therapies once thought impossible. "What if the money isn't there?" he asked. The Senate Budget Committee, which sets the limits within which funding committees must draft their spending bills, last week approved a 1999 budget that includes an 11 per cent boost in NIH funding, to \$15.1 billion.

First 'European PhDs' awarded in biotech

[MUNICH] The European Association for Higher Education in Biotechnology (HEduBT) has awarded its first certificates. The association, which was set up by a group of independent European scientists four years ago with the approval of the European Commission, last week credited two young scientists with 'European PhDs'. Both recipients had received their original PhDs at Swiss universities.

Those registering for HEduBT accreditation must already have a PhD from a European university. To win a European PhD, they must complete further courses, typically including patenting and ethics.

Asteroid watchers 'will talk first, go public later'

[WASHINGTON] Asteroid experts meeting last week in Houston, Texas, agreed to form a committee to review data on any future asteroid threat before going public with the news. The committee is intended to avoid the confusion that surrounded a research team's recent announcement of a possible asteroid collision with Earth in 2028. Within a day, other scientists were able to discount that possibility using archived data (see *Nature* 392, 215; 1998).

The committee is being created at the request of the US space agency NASA. Details on its membership and how it will operate have not been decided. But Donald Yeomans of NASA's Jet Propulsion Laboratory in Pasadena, California, says that the first step "is to make all data on near-Earth objects available to the small community of [people who compute asteroid orbits] almost immediately upon receipt so that we can all chew on it simultaneously".

Congressmen back plan to raise science spend

[WASHINGTON] The US House of Representatives Science Committee approves President Bill Clinton's plan to boost spending for research, but has quibbles about some of the details, says its annual "views and estimates" report released last week. Although it supports a 4 per cent increase for research next year, the committee said that projected research spending in subsequent years was inadequate. Republicans continue to disapprove of increases for the commerce department's Advanced Technology Program, and for environmental technology now included as part of Clinton's Climate Change Technology Initiative.

Academician delivers snub to Boris Yeltsin

[MOSCOW] Andrei Trofimuk, one of Russia's leading geologists and a co-founder of the Siberian branch of the Russian Academy of Sciences, has rejected the offer of a government award for 'services to the country'. The award is accompanied by a lifetime income of more than 10 times Russia's minimum salary.

Trofimuk, an opponent of the policies of President Boris Yeltsin, said he could not accept an award from someone he blames for "destroying the Soviet Union and launching the present reforms, [which are] disastrous". But the 86-year-old academician is not short of honours. He received six Order of Lenin awards — the highest honour in the former Soviet Union — two Stalin prizes and the title 'Hero of the Soviet Union'.

Patents, ownership and sovereignty

Sir — The debate about the draft European Commission directive on the legal protection of biotechnological inventions, soon to be submitted to the European Parliament (*Nature* **390**, 429; 1997), appears to have paid little regard to the views of laboratory bench scientists.

Results of a survey of users of the MRC-funded Human Genome Mapping Project (HGMP) Resource Centre (P. Glasner *et al.* *Genome News* Winter 26–32; 1997) provide a significant insight into the attitudes of researchers working on the HGMP on the wider issues surrounding genome research such as patenting and commercialization (see table). A stratified sample of 1,000 users generated a 52.5% response rate, with nearly 90% located in the United Kingdom. Two-thirds of our respondents were situated in either a hospital or higher education environment, and a similar proportion were research scientists; very few were in the private sector. More than 80% were actively involved in projects relevant to human genome research.

The greatest unanimity of view was expressed in a concern about the patenting of sequenced material. One respondent put the position starkly: “The patenting of any discovered genetic sequences is immoral and unethical. It should remain or be illegal.”

More than 85% of respondents either wholly or partially agreed that patenting could impede the development of diagnostics and therapeutics. More moderate responses, however, acknowledged that genome research is inextricably linked to financial interests, and that the issues arising from this need to be addressed.

A question on whether the current attempts to commercialize the results of the HGMP are premature did not arouse the same depth of feeling. Nearly one-third of respondents felt strongly that they are, another third mildly agreed and one respondent reiterated the need by all for access to knowledge in this area. This apparent paradox suggests the need for further attitudinal research.

Peter Glasner

Harry Rothman

*Science & Technology Policy Unit,
University of the West of England,
Bristol BS16 1QY, UK
e-mail: h-rothman@wpg.uwe.ac.uk*

Sir — You reported a call for a moratorium on patents on seeds held by the Consultative Group on International Agricultural Research (CGIAR) (*Nature* **391**, 728; 1998). This is ‘NGO-speak’. Samples are not being ‘patented’ but placed

under Plant Variety Rights (PVR) legislation. Although CGIAR claims ‘global trustee’ status for its collections, this is self-assigned; no country has ever formally allocated trustee status over its varieties to the CGIAR, and the legal status of the CGIAR collections remains uncertain.

Indeed, most of the varieties in CGIAR genebanks are duplicates of national collections. CGIAR collecting expeditions over the past three decades left original samples in national genebanks. These original samples have clearly always been under national sovereignty, and are certainly now covered by the articles of the Convention on Biodiversity.

At least one CGIAR institute — the International Center for Agricultural Research in the Dry Areas (ICARDA) in Syria — has recognized this. As each CGIAR institute is responsible for the terms of its material transfer agreement restricting the use of collections, ICARDA has sensibly asked those receiving samples to contact the country of origin if the recipient wishes to process samples for PVR. Attempts by any outside body to prevent this would be a clear breach of the biodiversity convention, which encourages such deals to support the use and conservation of biodiversity.

The World Bank, as the parent body of CGIAR, seems to be infringing the sovereignty of the country of origin by ignoring the provisions of the convention, and is even claiming retroactivity for a contractual agreement about CGIAR collections between CGIAR institutes and the Food and Agriculture Organization (FAO); this allows CGIAR monopoly rights over samples, and even over the interests of the country of origin. The confusion stems from the faulty agreement between FAO and CGIAR. This should now be changed to recognize national sovereignty over the collections that CGIAR has ‘borrowed’ from countries of origin.

David Wood

*13 Herons Quay,
Milnthorpe, Cumbria LA7 7HW, UK
e-mail: 113077.3244@compuserve.com*

Table 1 Views of users of the Human Genome Mapping Project Resource Centre*

	Disagree strongly	Disagree mildly	Indifferent	Agree mildly	Agree strongly
HGMP has drawn funds away from more worthwhile biological research, <i>n</i> = 507	41.0	26.8	24.1	6.3	1.8
HGMP poses new ethical issues for society, <i>n</i> = 509	9.8	13.6	18.9	34.8	23.0
Patenting of partial and uncharacterized cDNA sequences without knowledge of their biological function will impede the future development of diagnostics and therapeutics, <i>n</i> = 508	3.9	3.1	6.1	22.4	64.4
Current attempts to commercialize the results of the HGMP are premature, <i>n</i> = 503	2.6	8.3	27.8	32.0	29.2
HGMP will advance technology and techniques more than biological theory and science, <i>n</i> = 502	11.4	24.9	26.9	28.1	8.8

*Results are percentages.

Send not to know...

Sir — Jonathan Slack was perceptive in his book review where he discussed the challenges faced by students of developmental biology (*Nature* **391**, 857–858; 1998). One problem of terminology causes particular confusion. The example from Lewis Wolpert’s book *Principles of Development* illustrates the

point well, which Slack did not address directly.

The example was: “As a result of Toll receptor activation cactus protein is degraded....” This sentence is about the protein Toll, which is a receptor that is activated by binding its ligand, Spaetzle. In other words Toll is the receptor for Spaetzle, just as the insulin receptor is the receptor for insulin. The sentence does not

refer to a receptor for which Toll is the ligand. Such confusing use can be found in many textbooks on developmental biology where it causes great difficulty for students new to the subject. Is it too late for clarity on this point?

Robert W. Old

*Developmental Biology Group,
University of Warwick,
Coventry CV4 7AL, UK*

Lifting the veil on perverse subsidies

Many government subsidies serve useful purposes, but others adversely affect the economy and the environment. A forthcoming report suggests that governments could profit from scrapping such 'perverse' subsidies.

Norman Myers

When the leaders of the eight leading industrialized nations meet in Birmingham, England, in May for their annual G8 summit, they might ponder an issue that has been given little attention in the councils of power but is an important factor in the global economy. It is the issue of 'perverse' subsidies — subsidies that are adverse in the long run to both the economy and the environment.

A prime example currently in the news is the German government's support for coal-mining. So huge are the subsidies, DM12.2 billion (US\$6.7 billion) a year, or DM142,000 per miner per year, that it would be economically more efficient to close all the mines and send the workers home on full pay for the rest of their lives. There would be benefits too for the environment: less coal pollution, such as acid rain and global warming.

Perverse subsidies are prominent in five leading sectors: agriculture, fossil fuels and nuclear energy, road transport, water and fisheries. Subsidies for agriculture foster overloading of croplands, leading to erosion of topsoil, pollution from synthetic fertilizers and pesticides, and release of greenhouse gases. Subsidies for fossil fuels aggravate pollution effects of acid rain, smog and global warming. Subsidies for road transport promote pollution at local, national and global levels, including excessive road building and loss of landscape. Subsidies for water encourage misuse and overuse of supplies that are increasingly scarce. And subsidies for fisheries support overharvesting of depleted fish stocks. Not only do these environmental ills have economic costs, but the subsidies hinder the efficiency of economies overall¹⁻³.

This is not to say that subsidies cannot serve many useful purposes. They can overcome deficiencies of the marketplace, support disadvantaged sections of society and promote environmentally friendly technologies. We sometimes need a bit of positive distortion, otherwise we might never get as much as we want of, for example, non-polluting and renewable sources of energy (particularly when fossil fuels with their many ills are often subsidized ten times more). The same applies to recycling, agricultural set-asides and other subsidies beneficial to the environment and the economy alike.

Hidden subsidies

Subsidies come in many shapes and sizes. They range from financial transfers to opportunity costs, and can be direct or indi-



Incentive to overfish: the global fisheries catch is subsidized to the tune of \$22 billion a year.

rect, overt or covert. In addition to subsidies of conventional and formal type, there is a host of implicit subsidies, especially in the form of environmental externalities. For example, car drivers pollute everyone's atmosphere without compensating everyone, so they effectively gain a benefit at everyone's expense. Much the same applies when farmers spray pesticides which then extend their toxic effects into everyone's ecosystems. These activities should count as implicit subsidies in both spirit and substance, even though they are not dispensed by a government department with appropriate paperwork. They are economically distorting and socially inequitable, as well as environmentally harmful.

Environmental externalities are widespread, important and growing fast. Global soil erosion levies unintended costs on society of about \$150 billion per year⁴, and pesticides harm society's interests to the extent of at least \$100 billion per year⁵; these two items alone mean that such hidden subsidies are almost as large as the formal subsidies in agriculture. In Costa Rica, the depletion of soils, forests and fisheries results in a reduction of 25–30% in potential economic growth⁶.

Despite their importance and the huge literature on them, formal subsidies, let alone implicit ones, are often difficult to document. What little information is available tends to be incomplete, imprecise and inconsistent. Governments are reluctant to admit that they hand out subsidies of myriad sorts in munificent amounts. Still less do they want to concede that some could be ill-conceived, out of date, politically dubious or off target. And in many instances governments do not compile records on such contentious issues. So data are often patchy, suggesting that many estimates are underestimates.

If it is hard to assess subsidies overall, it is still more difficult to estimate perverse subsidies comprehensively. But here I present the findings of a forthcoming report¹ that attempts to do just that. The findings are to be viewed as conservative and cautious. But it is worthwhile trying to provide rough estimates so that political leaders, policy-makers and the public can be appraised of the scale of these damaging subsidies.

The scale of the problem

Table 1 shows that subsidies in five leading sectors are estimated to total around \$1,900 billion per year, and perverse subsidies \$1,450 billion. Plainly, perverse subsidies can exert a highly distorting effect on the global economy of \$29 trillion and greatly harm our environment. On both counts, they foster unsustainable development. Ironically the total of almost \$1.5 trillion is two-and-a-half times the Rio Earth summit's budget for sustainable development, a sum governments claimed could not be found.

The OECD (Organization for Economic Cooperation and Development) countries account for two-thirds of all subsidies and an even larger share of perverse subsidies; the United States accounts for more than a fifth of perverse subsidies; and road transport alone accounts for 48% of all subsidies and 44% of perverse subsidies.

The total for perverse subsidies may seem unduly large, but remember that many environmental externalities (including global warming, which could prove to be as big as all the rest put together) are either underestimated or omitted owing to lack of documentation of economic costs. In just the road transport sector, environmental externalities could account for \$1 trillion worldwide⁷ out of total sector costs of \$2 trillion a year⁸.

So what are the leading instances of these

Table 1 Subsidies (billions of dollars per year)

Sector	Conventional subsidy	Environmental externality	Total (range)	Perverse subsidies (range)
Agriculture	325	250	575	460 (390–520)
Fossil fuels and nuclear energy	145	–	145	110
Road transport	558	359	917 (798–1,041)	639
Water	60	175	235	220
Fisheries	22	–	22	22
Total	1,110	785	1,895	1,450

Conventional subsidies include established and readily recognized subsidies, including both direct financial transfers and indirect supports such as tax credits. Some of the estimates are supported by ranges (see text). In some instances, ranges are not included owing to insufficient agreement. Data are too patchy and disparate to allow even a reasonably agreed value of the environmental externalities for fossil fuels and nuclear energy.

perverse subsidies? First, the global ocean fisheries catch — well above sustainable yield — is annually worth more than \$100 billion at the dockside, whereupon it is sold for some \$80 billion, the shortfall made up with government subsidies. This results in the commercial extinction of major fisheries, as well as the bankruptcy of fishing businesses and much unemployment⁹. Second, one US government agency heavily subsidizes irrigation for crops that another agency pays farmers not to grow.

Third, also in the United States, petrol is now cheaper than bottled water, and several other aspects of US road transport are similarly priced artificially low, all thanks to extensive subsidies. The unpaid costs of road transport are at least \$464 billion per year, or \$1,700 per head of population. Hidden subsidies for oil by default create an energy policy that is the opposite of the government's stated priorities: for instance, they prolong the risky dependence on foreign supplies, especially from the Persian Gulf, and discourage investment in new, cleaner technologies such as hyper-cars and other revolutionary forms of energy efficiency¹⁰.

The annual total of perverse subsidies is larger than all but the five largest national economies, and is larger than the top 12 corporations' annual sales. It is more than twice global military spending per year, and three times the annual cash income of the 1.3 billion poorest people.

Even worse are the indirect costs to economies. Perverse subsidies push up the costs of government, inducing higher taxes (higher prices as well), and aggravating governments' budget deficits; divert government funds from better opportunities for fiscal support; undermine investment decisions and reduce pressure for businesses to become more efficient; tend to benefit the few at the expense of the many, and the rich at the expense of the poor; and foster environmental degradation in many forms, which, apart from their intrinsic harm, act as a further brake on economies.

If perverse subsidies were reduced, there would be a double dividend. First, there would be an end to the obstacles they impose on sustainable development. Second, there

would be a huge stock of funds — on a scale unlikely to become available through any other source — for sustainable development. In the case of the United States, for instance, they would amount to more than \$300 billion, or more than the Pentagon budget and two-and-a-half times the federal deficit. If just half of the perverse subsidies were to be phased out, just half of the funds released would enable most governments to abolish their budget deficits at a stroke, to reorder their fiscal priorities fundamentally, and to restore environments more vigorously than through any other single measure.

Finding a solution

The political climate for the reform of perverse subsidies is probably better than it has been for decades. Many governments are espousing the marketplace economy with its reduced government intervention; and many governments face fiscal constraints that give them further incentive to reduce subsidies. This applies particularly to the 'transition economies'. Fortunately, several governments have greatly reduced or even abolished some of their perverse subsidies, thus supplying a helpful precedent.

New Zealand has eliminated virtually all its agricultural subsidies since the early 1980s, even though — or perhaps because — its economy is more dependent on agriculture than that of most OECD countries. Today there are more farmers in New Zealand than when the subsidy phase-out began. Several Latin American countries, notably Chile and Argentina, have recently cut their agricultural subsidies.

Russia has reduced its fossil-fuel subsidies from \$29 billion in 1990–91 to \$9 billion in 1995–96; China has cut its subsidy from \$25 billion to \$10 billion; and similar drastic reductions have been made in India, Poland and Belgium¹¹.

Since the mid-1980s, Bangladesh and several other Asian countries have recognized that excessive application of nitrogenous fertilizers, encouraged by extravagant subsidies, are economically wasteful (decline in marginal benefits) and highly polluting (eutrophication of waterways and threats to supplies of drinking water). Pakistan has

reduced its fertilizer subsidies from \$178 million to \$2 million per year, Bangladesh from \$56 million to zero, and the Philippines from \$48 million to zero¹².

Policy openings abound. Governments can ride the strong tide of 'marketplace-ism'. Resistance to subsidies also stems from the growing privatization ethos. India's subsidies total more than 14% of its gross domestic product (GDP), yet the government wishes to reduce its fiscal deficit to under 4% of GDP, providing motivation to cut subsidies drastically.

These formidable opportunities are matched by formidable obstacles. Special-interest groups often feel so addicted to their 'entitlements' that they recoil from talk of cutting perverse subsidies. They find allies in bureaucratic roadblocks and institutional inertia. There is uncertainty about how the reduction of perverse subsidies, however rational in principle, will work out in practice; for instance, will it mean that businesses lose out to competitors abroad?

There are various ways of overcoming these obstacles. Governments could formulate alternative policies that target the same subsidy objectives better, while also compensating losers; develop a policy strategy that encourages subsidy removal by, say, reducing government controls generally and freeing up the economy to marketplace discipline; or introduce 'sunset' provisions that require surviving subsidies to be rejustified periodically, thus avoiding entrenchment.

All these measures can be reinforced by making perverse subsidies more transparent, especially their economic and environmental effects and their costs to taxpayers and consumers¹³. A US citizen pays taxes of at least \$2,000 a year to fund perverse subsidies, and pays almost another \$2,000 through increased costs for consumer goods and through environmental degradation. □

Norman Myers is at Green College, Woodstock Road, Oxford OX2 6HG, UK.

- Myers, N. & Kent, J. *Perverse Subsidies* (International Institute for Sustainable Development, Winnipeg, 1 May 1998).
- de Moor, A. *Perverse Incentives: Hundreds of Billions of Dollars in Subsidies Now Harm the Economy, the Environment, Equity and Trade* (Earth Council, San Jose, Costa Rica, 1997).
- Roodman, D. M. *Paying the Piper: Subsidies, Politics and the Environment* (Worldwatch Institute, Washington DC, 1996).
- Pimentel, D. et al. *Science* **267**, 1117–1122 (1995).
- Pimentel, D. (ed.) *Techniques for Reducing Pesticide Use: Economic and Environmental Benefits* (Wiley, Chichester, 1997).
- Cruz, C. et al. *Population Growth, Poverty and Environmental Stress: Frontier Migration in the Philippines and Costa Rica* (World Resources Institute, Washington DC, 1992).
- von Weizsacker, E. U. et al. *Factor Four: Doubling Wealth, Halving Resource Use* (Earthscan, London, 1997).
- Litman, T. *Transportation Cost Analysis* (Victoria Transport Policy Institute, Victoria, Canada, 1996).
- Safina, C. *Song for the Blue Ocean* (Henry Holt, New York, 1998).
- Lovins, A. B. *Energy Policy* (April 1996).
- Rajkumar, A. S. *Energy Subsidies* (World Bank, Washington DC, 1996).
- World Bank *Expanding the Measure of Wealth* (World Bank, Washington DC, 1997).
- Gale, R. J. P. & Barg, S. R. (eds) *Green Budget Reform* (Earthscan, London, 1995).

Acknowledgements. I thank the MacArthur Foundation, Chicago, for research support.

Giant submarine landslides

Euan G. Nisbet and David J. W. Piper

Submarine landslides can generate enormous turbidity currents that carry 500 km³ or more of sediment down to the oceans' abyssal plains. The slides can cause tsunamis, and may release large amounts of methane to the air.

Late in 1929, many banks failed. Most of them were on Wall Street, but one was already under water, off Newfoundland. On 18 November of that year, the Grand Banks earthquake¹ set off a 20-km³ submarine landslide from which 27 people died. The slide in turn generated an erosive, turbulent sediment flow, or turbidity current, that carried 200 km³ of debris into deep water. Trans-Atlantic submarine telegraph cables were cut, which did not help in resolving the economic problems. The times when a series of the cables were cut gave the flow's speed, as great as 65 k.p.h. Submarine landslides such as this are not uncommon, and they were particularly frequent during the late glacial period, 15,000 to 25,000 years ago, when sea level was lower².

These landslides can spawn dense, turbulent, sediment-laden flows that speed down the continental slope to deposit 'turbidites' on the abyssal plains (Fig. 1). Turbidite sequences can give high-resolution palaeo-environmental records³, and on page 377 of this issue, Rothwell *et al.*⁴ report the discovery of just such a record. They have found a 'megaturbidite', 8–10 m thick, across the deep floor of the western Mediterranean.

The volume of the deposit, 500 km³, is enough to cover all of Texas waist deep in

mud and sand. The source of the material would have been a submarine landslide on the continental slope seaward of a large river draining glaciated Europe (such as the Rhône or Ebro). At the time of the slide, about 22,000 years ago, ice accumulation on land had lowered sea level significantly. The event may have been dramatic: an immense submarine flow, followed by huge waves crashing onto the Mediterranean coasts. Rothwell *et al.* have not identified the source of the sediment. But could this be where some of the glacial detritus from the Alps or the Pyrenees ultimately went? Could fragments from the chiselled face of the Matterhorn be on the floor of the Mediterranean?

Submarine slides occur especially where huge piles of unstable sediment build up (and also off volcanic islands; see box, overleaf). In some cases, deposits of methane hydrate — ice-like solids formed of water and gas — are implicated in causing them. Decay of organic matter in sediments produces gas, increasing pore pressure and reducing sediment strength. Under ambient pressures and temperatures, methane hydrate grows in the upper layers of sediment on the continental slope; below, free gas collects in sediment. The base of these hydrate layers can be identified by reflection

seismology, forming a bottom-simulating reflector, which commonly cross-cuts reflections from the sediment strata.

During glaciation, as sea level drops, the confining pressure in the sediment decreases and the lower hydrate layers break down, releasing free gas that is trapped in the sediment pile. The immediate trigger of a slide may be a small earthquake, new sediment brought down by a flooding river, or even a large storm. Once movement is initiated, gas pools in sediment below the hydrate may rupture. Rothwell *et al.* speculate that hydrate breakdown was involved in the creation of the megaturbidite they have discovered, and most large slides do indeed seem to be linked with hydrate destabilization, or 'glide' on hydrate layers. The case cannot, however, be proved until a smoking methane release is found.

Slides can be enormous. Off the mouth of the Amazon, slides capable of producing megaturbidites as big as that described by Rothwell *et al.* took place in late glacial time⁵. One has left a 120-m headscarp, a submarine cliff as high as a 40-storey building. Failures of this size appear to be regular: buried slides of similar size dating from 45,000 years ago have also been found. In some, the evidence for methane hydrate involvement is much stronger than in the case described by Rothwell *et al.* For example, the late glacial Cape Fear slide, off the Carolinas on the east coast of the United States, has left a 50-km-long scar in sediment over a gas hydrate field².

One of the best-studied sets of submarine landslides is the Storegga ('Great Edge') suite, three slides whose immense headwall, nearly 300 km long, runs roughly along the edge of the continental shelf off the coast of Norway⁶. Debris, up to 450 m thick, is spread a distance of 800 km out to the 3.6-km-deep abyssal floor.

The enormous first slide occurred 30,000–50,000 years ago, and was much larger than that described by Rothwell *et al.* — the volume involved would cover the area of Alaska above head height, and has left a scar much bigger than Maryland. In the second slide (6,000–8,000 years ago), two slabs of up to 10 × 30 km were moved 200 km downslope. On the Scottish coast, a tsunami from this slide deposited sediment up to 4 m above high-water mark⁷ (though the speculation that it swept a monster into Loch Ness must, alas, be dismissed). The glide plane of the first Storegga slide is at the same depth as the base of a gas hydrate layer, seen as a bottom-simulating reflector which cross-cuts bedding reflections, on the margin of the slide scar. Precisely dating the event is difficult. But the disruption may have smothered nearby deep-sea coral reefs (J. Wilson, personal communication), which may give an age.

What of the possible climate connection through methane release, which is potential-

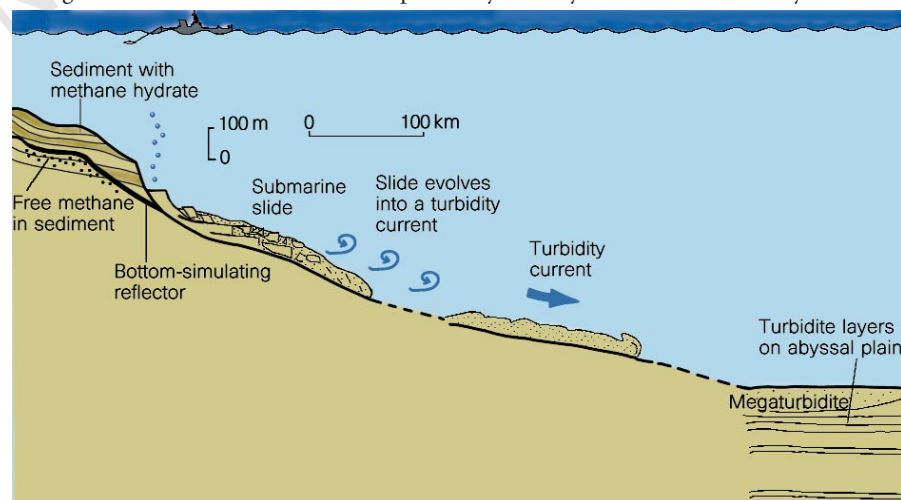


Figure 1 The likely mode of formation of a megaturbidite deposit, such as that described by Rothwell *et al.*⁴. Unstable sediment accumulations collapse when perturbed, maybe with associated release of methane, resulting in a submarine landslide and flow of dense currents of sediment (turbidity currents) down a continental slope. The end result is turbidite sequences on the abyssal plain.

Slides from volcanic islands

Some of the most spectacular submarine landslides are from volcanic islands, where volcanoes have built up vast edifices of lavas onto weak substrates of oceanic sediments. These fail and cause the flank of the volcanic structure to collapse into the deep. This process seems to be a normal stage in the evolution of the Hawaiian islands¹², which are surrounded by vast aprons of debris of over 100,000 km² in extent. Slides can be over 200 km long, with



volumes of about 5,000 km³: the sonograph shown here is about 150 km across, and is of avalanches north of the islands of Oahu and

Molokai¹². Some of the Hawaiian slides may have been rapid, creating immense tsunamis which have left marine deposits at altitudes of well over 300 m above sea level on Lanai, one of the smaller Hawaiian islands¹², and which reached as far as Australia. Similar collapses of volcanic edifices have also taken place off the Canary Islands, generating megaturbidites¹³, and off Réunion Island in the Indian Ocean¹⁴.

E.G.N. & D.J.W.P.

ly large, especially if sediment rich in free gas is ruptured? Some giant slides in late glacial time could, individually, have released many times the amount of methane in the air at that time (roughly 1 gigatonne). A slide as big as the first Storegga collapse could have released 5 gigatonnes of methane or more⁸. If it did, for a few years the greenhouse impact could have been profound⁹.

At certain times in the glacial cycle, a large slide could have triggered sudden warming, melting ice and raising sea level. Following emission from a large slide, warming and consequent transgression over the Arctic could have induced smaller but more frequent releases of gas from different, permafrost, hydrate sources. This, together with methane from swamps, could have helped sustain increased methane levels over longer periods¹⁰. Methane peaks from slide bursts would be short-lived, but should be just detectable in ice-core methane records⁹. But the evidence for huge, sudden methane bursts is indirect: their confirmation or denial is a challenge for future ice core work¹¹.

Finally, there is the question of whether catastrophic sediment failures could occur today. Although sediment failures were commoner when sea level was close to its glacial low (subsequent sea-level rise may have had a stabilizing effect), huge piles of sediment have since accumulated in many places near steep slopes. It is likely that the landslide described by Rothwell *et al.* occurred off the French coast, and indeed one such smaller event happened there about 20 years ago. In 1979 a 0.15-km³ slide off Nice airport caused a tsunami that killed 11 people.

The hazards of submarine landslides range from the economic impact of severed fibre-optic submarine cables and pipelines, to the devastating effects of massive

tsunamis or involvement in the slide itself. Small slides near densely populated areas may be particularly dangerous, so that cities close to deltas (Vancouver, for instance) may be most at risk. Although large events are

Human physiology

Life without leptin

Stephen O'Rahilly

The detailed study of nature's experiments on humans has an impressive track record in the field of physiology. In previous centuries such studies usually required a bizarre injury to expose the previously concealed organs of an unfortunate patient to the eyes of a curious physician. But in the late twentieth century we can observe illuminating accidents that occur at the level of DNA. The power of this approach is exemplified by two studies of the fat-derived hormone leptin, one by Clément *et al.*¹ reported on page 398 of this issue, and the other by Strobel *et al.*² in *Nature Genetics*. By studying adult humans who are genetically deficient in leptin signalling, the authors show that leptin is needed for the initiation of puberty and establishment of secondary sexual characteristics. These studies also support the idea that leptin is central to the control of body fat mass in man, and provide more information about other neuroendocrine and metabolic sequelae of human leptin deficiency.

The discovery that fat cells secrete a circulating hormone, leptin, which interacts with receptors in the hypothalamus to regulate appetite and energy expenditure came from the study, by Friedman and colleagues³, of the genetic basis for severe obesity in mice with mutations in the *obese* gene (*ob/ob*

rare, the far-field danger of tsunamis is widespread. It would be worth being better informed. □

Euan G. Nisbet is in the Department of Geology, Royal Holloway College, University of London, Egham TW20 0EX, UK.

David J. W. Piper is in the Geological Survey of Canada, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada B2Y 4A2.

1. Heezen, B. C. & Ewing, M. *Am. J. Sci.* **250**, 849–873 (1952).
2. Paull, C. K., Buelow, W. J., Ussler, W. III & Borowski, W. S. *Geology* **24**, 143–146 (1996).
3. Maslin, M. A., Burns, S., Erlenkeuser, H. & Hohnemann, C. *Proc. Ocean Drilling Prog., Sci. Res.* **155**, 305–318 (1997).
4. Rothwell, R. G., Thomson, J. & Kähler, G. *Nature* **392**, 377–380 (1998).
5. Piper, D. J. W. *et al. Proc. Ocean Drilling Prog., Sci. Res.* **155**, 109–146 (1997).
6. Bugge, T., Belderson, R. H. & Kenyon, N. H. *Phil. Trans. R. Soc. Lond. A* **325**, 357–388 (1988).
7. Long, D. E., Smith, D. E. & Dawson, A. G. *J. Quat. Sci.* **4**, 61–66 (1989).
8. Paull, C. K., Ussler, W. III & Dillon, W. P. *Geophys. Res. Lett.* **18**, 432–434 (1991).
9. Thorpe, R. B., Law, K. S., Bekki, S., Pyle, J. A. & Nisbet, E. G. *J. Geophys. Res.* **101**, 28627–28635 (1996).
10. Nisbet, E. G. *J. Geophys. Res.* **97**, 12859–12867 (1992).
11. Brook, E. J., Sowers, T. & Orchard, J. *Science* **273**, 1087–1091 (1996).
12. Moore, J. G. *et al. J. Geophys. Res.* **94**, 17465–17484 (1989).
13. Masson, D. G. *Geology* **24**, 231–234 (1996).
14. Ollier, G., Cochon, P., Lénat, J. F. & Labazuy, P. *Sedimentology* (in the press).

mice). As well as being severely hyperphagic — that is, having markedly increased appetite — these mice have a low body temperature. So, abnormalities of both energy intake and energy output combine to produce their explosive weight gain. A wider role for leptin in mammalian physiology is suggested by other features of the *ob/ob* mouse. Homozygous mutant animals are short and infertile, with abnormally high circulating levels of glucose, insulin and cortisol. All of these features can be reversed by giving the mice leptin⁴.

Leptin is also thought to be involved in control of the hypothalamic–pituitary–gonadal (HPG) axis. Body fat has been linked with the initiation of increased activity of the HPG axis at puberty, and its maintenance in later life⁵. As well as restoring fertility in mutant *ob/ob* mice⁶, leptin administration accelerates the onset of reproductive function in normal rodents^{7,8}. Thus, plasma leptin may act as a signal of nutritional state — it will allow pubertal growth and sexual development only when there are adequate nutritional stores to sustain it. In a similar vein, caloric restriction in adult animals (including man) rapidly leads to shutdown of the HPG axis. Teleologically, it makes sense to avoid becoming pregnant when there are inadequate nutritional stores to

support a growing fetus. Ahima *et al.*⁹ have shown that, in starved mice, plasma leptin falls rapidly and this is accompanied by a range of neuroendocrine changes, including the switching off of reproductive function. When leptin is injected into these mice, the neuroendocrine and reproductive effects of starvation are blunted. So leptin could be the signal that is used by the brain to modulate reproductive function according to nutritional state.

Clément *et al.*¹ and Strobel *et al.*² now provide compelling evidence that leptin is required for progression through puberty in man. Congenital leptin deficiency was first reported¹⁰ in two young children, so this question could not be addressed. Clément *et al.* describe a large consanguineous family of Kabilian origin, in which three morbidly obese sisters are homozygous for a splice-site mutation in the leptin receptor. This mutation leads to a truncated form of the receptor with no signalling function, and it affects all isoforms of the receptor. This is the first description of a human genetic defect at this locus, although the leptin receptor is known to be mutated in the *db/db* mouse and the *fa/fa* rat (that is, rodents with homozygous mutations in the *diabetes* and *fatty* genes). In two of the affected sisters (ages 13.5 and 19 years) there were no signs of pubertal development, and the endocrine measurements were again consistent with a hypothalamic defect.

Strobel *et al.* describe a Turkish family in which a homozygous missense mutation in the leptin gene results in low plasma leptin and morbid obesity in three affected members, two of whom are adults. The 22-year-old male has clearly failed to go through puberty, and endocrine measurements indicate impaired hypothalamic activation of the pituitary gonadotroph. Little information is available on the 34-year-old female, other than that she has never menstruated.

Does leptin merely provide a tonic, 'background' signal on which other initiators act, or is it more actively involved in the initiation of puberty? Most cross-sectional studies of children indicate that changes in plasma leptin levels follow changes in fat mass with age¹¹. One study¹² has suggested that a brief 'pulse' of leptin may precede the onset of puberty, at least in males. Studies of leptin therapy in leptin-deficient patients will be illuminating in this regard.

Is leptin necessary for function of the HPG axis at all stages of life? In the male fetus and early neonate, androgens are secreted from the testis in a gonadotropin-dependent manner. Evidence for this includes the frequent presence of a micropenis in male neonates with congenital hypopituitarism, or inherited abnormalities in secretion of gonadotropin-releasing hormone¹³. However, as neither of the males studied by Clément *et al.* or Strobel *et al.* (nor a third

affected male; unpublished observations of my own group) had a neonatal micropenis, this raises the possibility that gonadotropin release in early life may be independent of leptin.

Thus, the absolute requirement for leptin in controlling body fat mass and regulating reproduction is firmly shared between mouse and man. But other consequences of absent leptin signalling seem to differ between the species. In particular, the high levels of cortisol seen in the *ob/ob* and *db/db* mice are not found in any of the humans studied. Given the adverse metabolic consequences of glucocorticoid excess, this might explain the finding that the human subjects are, in general, also much less severely hyperinsulinaemic and hypertriglyceridaemic. Moreover, only one of the eight subjects in the three families has, thus far, developed diabetes mellitus. The mice also have stunted linear growth, whereas only the homozygous members of the family with the leptin-receptor mutation show any indication of growth retardation — and this is mild. Finally, unlike the mice, affected humans have a normal body temperature, and their resting metabolic rates are not obviously low. Thus, the obesity in leptin-deficient humans may be more the result of excess energy intake, although more detailed study of energy expenditure is required before any definitive statements can be made.

The description of human life without leptin not only provides important lessons in

human and comparative physiology, but the studies have also identified monogenic forms of severe obesity in man. Although rare, these forms are potentially amenable to treatment — either with leptin or with cytokines that might mimic the action of leptin in the hypothalamus¹⁴. The growing evidence that humans can be genetically hardwired to become severely obese should eventually lead to a more widespread realization that morbid obesity is a disease requiring further scientific research, rather than a failure of will-power requiring sanctimonious moral opprobrium. □

Stephen O'Rahilly is in the Departments of Medicine and Clinical Biochemistry, University of Cambridge, Hills Road, Cambridge CB2 2QQ, UK. e-mail: sorahill@hgmpr.mrc.ac.uk

1. Clément, K. *et al.* *Nature* **392**, 398–401 (1998).
2. Strobel, A., Issad, T., Camoin, L., Ozata, M. & Strosberg, A. D. *Nature Genet.* **18**, 213–215 (1998).
3. Zhang, Y. *et al.* *Nature* **372**, 425–432 (1994).
4. Campfield, L. A., Smith, F. J., Guisez, Y., Devos, R. & Burn, P. *Science* **269**, 546–549 (1995).
5. Cameron, J. L. *Nutr. Revs* **54**, s17–s22 (1996).
6. Chehab, F. F., Lim, M. E. & Lu, R. *Nature Genet.* **12**, 318–320 (1996).
7. Chehab, F. F., Mounzih, K., Lu, R. H. & Lim, M. E. *Science* **275**, 88–90 (1997).
8. Ahima, R. S., Dushay, J., Flier, S. N., Prabakaran, D. & Flier, J. S. *J. Clin. Invest.* **99**, 391–395 (1997).
9. Ahima, R. S. *et al.* *Nature* **382**, 250–252 (1996).
10. Montague, C. T. *et al.* *Nature* **387**, 903–907 (1997).
11. Clayton, P. E. *et al.* *Clin. Endocrinol.* **46**, 727–733 (1997).
12. Mantzoros, C. S., Flier, J. S. & Rogol, A. D. *J. Clin. Endocrinol. Metab.* **82**, 1066–1070 (1997).
13. Aaronson, I. A. *J. Urol.* **152**, 4–14 (1996).
14. Gloaguen, I. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 6456–6461 (1997).

Liquid crystals

Out of the groove

Jos van Haaren

Cellular phones, computer games, electronic diaries and notebook computers are common products nowadays, and they all rely on liquid-crystal displays (LCDs). In turn, high-quality displays depend on precise orientation control of their liquid-crystal materials. How is this done? A popular simplified explanation is that grooves in the supporting substrates dictate liquid-crystal alignment. This does not actually happen in commercial displays; but Bryan-Brown and co-workers (on page 365 of this issue¹) remove the discrepancy between the simple picture and the reality. They have made an LCD with lithographic grooves and a new way of electrically manipulating liquid-crystal orientation. Without increasing production costs by much, this could improve the optical performance of LCDs to the point at which they could be used in desktop monitors and televisions.

Liquid crystals consist of elongated, organic molecules, with an average orientation that specifies the local direction of the medium. They form an intermediate phase

between disordered liquids and well-organized crystals. Some properties are shared with liquids, such as the ability to completely fill the gap between a display's glass substrates, and the easy rotation of the molecules under the influence of external electrical fields. Other aspects of liquid crystals are solid-like, such as optical and dielectric anisotropy, and energy in elastic deformations.

The dual nature of the phase is exploited in LCDs, but it also causes problems. The liquid crystal needs to have a single orientation in the display cell, which consists of a cavity typically 5 µm high and up to a magazine page in area. But, having been forced to flow into this cavity via holes in the peripheral seal, the crystal doesn't necessarily form a single domain. In practice, an orientation is imposed using a surface treatment of the substrates (today, a thin polymer layer), which is aligned by rubbing with a velvet-like cloth. This process² goes back to 1925.

Reliable liquid-crystal orientation is not only of practical relevance; it is also an intriguing phenomenon, studied by many

scientists with experimental and theoretical backgrounds. The effect of rubbing is often depicted as making small grooves in surfaces. This is a popular image, partly because of elegant theoretical work by Berreman³ in 1972 explaining liquid-crystal orientation on grooved substrates. But the presence of grooves on real substrates has not been confirmed by experiment; instead, stretching of polymer chains appears to be responsible for the orientation of the liquid crystal⁴.

So, how is orientation used to make displays? Most LCDs use a twisted nematic⁵ configuration (Fig. 1a). In the absence of an electric field, the liquid-crystal molecules have a planar orientation, with a 90° twist going from the back of a display to the front. There are crossed polarizers on either face of the display which, without the liquid crystal, would block the passage of light; but the 90° twist rotates the light, and so allows it to pass through. To turn the display black, the molecules are switched to a more upright orientation by an electric field perpendicular to the substrates. The rotation is disturbed, and so electronic data are translated into an image.

This twisted nematic switching mechanism is universal in today's eight-billion-dollar-a-year LCD industry. But it has a severe disadvantage: the brightness depends strongly on viewing angle. That is OK for a calculator, but it is no good for a TV screen.

An improvement was made three years ago by Hitachi. Their new in-plane-switching⁶ displays have a better viewing-angle performance than twisted nematic LCDs. The orientation of the liquid-crystal molecules is changed while keeping them parallel to the substrates (Fig. 1b). Hitachi and other companies have exhibited prototypes and pilot products, and are preparing for mass production. But these displays also have disadvantages: the driving electric fields have to be parallel to the substrates, which involves a relatively complicated structure of interdigitated electrodes. That will make them expensive, to begin with. Furthermore, they are dimmer than conventional displays, because the part of the LCD on top of the interdigitated electrodes is not effective; and it is difficult to make their response speed high enough for video applications.

Bryan-Brown and co-workers¹ may have found a way around these problems. Their upper substrate is grooved and treated with a surfactant to make the liquid-crystal molecules orientate themselves perpendicular to the surface. Using rubbed-polymer treatment, molecules at the lower substrate are orientated parallel to it and at 90° to the grooves in the grating above. In between these substrates, the molecules have a splayed, non-twisted orientation.

The liquid-crystal material in the new cells is of a type that tends to align perpendicular to an electric field. (This is different from the materials in the conventional displays

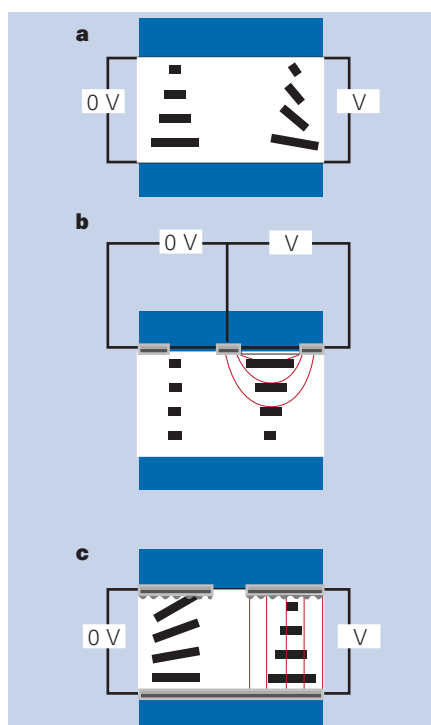


Figure 1 Lining up a liquid crystal: molecular orientations for three different switching modes of a liquid-crystal display. Twisted nematic effect (a), in-plane switching (b) and voltage-controlled twist (c), in which a voltage forces the molecules near the upper substrate to orientate themselves more in the plane of the substrate, and eventually to align along the grooves in a grating. This has the advantage of simple electrodes, as in a, and yet good viewing-angle response, as in b.

(Fig. 1a) which tend to align parallel to the electric field.) So applying a voltage reduces the tilt of the molecules near the grating. These molecules tend to relax the resulting mechanical stress by aligning parallel to the grooves. The molecules at the rubbed substrate maintain the original orientation, and the molecules in between form a smooth transition between the two preferential directions at the boundaries. The result is an electrically induced twist (Fig. 1c).

More work is needed before we can be sure that applications will follow, but this method could markedly improve the image quality of LCDs, without using complicated electrodes. An old idea in liquid-crystal science has led to a new answer to today's industrial challenges. □

Jos van Haaren is at Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands.

e-mail: vhaaren@natlab.research.philips.com

1. Bryan-Brown, G. P., Brown, C. V., Sage, I. C. & Hui, C. V. *Nature* **392**, 365–367 (1998).
2. Zocher, H. *Naturwissenschaften* **13**, 1015–1021 (1925).
3. Berreman, D. W. *Phys. Rev. Lett.* **28**, 1683–1686 (1972).
4. van Aarle, N. J. A. M. & Tol, A. J. W. *Macromolecules* **27**, 6520–6526 (1994).
5. Schadt, M. & Helfrich, W. *Appl. Phys. Lett.* **18**, 127–128 (1971).
6. Oh-e, M. & Kondo, K. *Appl. Phys. Lett.* **67**, 3895–3897 (1995).



100 YEARS AGO

Harvey delivered his first course of Lumleian Lectures in 1616. It was in these lectures that he first propounded his views on the circulation of the blood, and demonstrated the anatomical and experimental evidence on which his conclusions were based. These demonstrations were, as he tells us, annually repeated at the Lumleian Lectures for nine successive years. It was only after this long probation that Harvey ventured to give his discoveries to the world. This he did in the form of a small Latin quarto of 76 pages... This little book is in several respects a remarkable one. It constitutes the earliest record we possess of a really scientific investigation in the domain of biology based on systematic observation and experiment. Although written 270 years ago, the work is essentially modern in tone and method. It is, in fact, the precursor and prototype of the scientific "monograph" of our own day, and stands favourable comparison with the best master-pieces of recent times. In this treatise Harvey established absolutely the fact of the circulation of the blood, and the fact that the heart was the propulsive agent in the movement. But he was unable, from his want of a microscope, to indicate the precise path along which the blood travelled from the terminal arteries to the commencing veins. He erroneously conjectured that the blood percolated the organs and tissues as water percolates the earth and produces springs and rivulets.

From *Nature* 24 March 1898.

50 YEARS AGO

Objectivity was never Malinowski's method. Like all great teachers, he was prone to generalize recklessly from a special instance in order to drive home a point of principle. ... The social scientist can never hope to be entirely detached, and it is arguable how far he can, or should, avoid imposing his own value-judgments upon the material of his study. My own feelings, in the present instance, are that it would need only a slight recasting of Malinowski's argument to demonstrate that totalitarian Russia rather than democratic America is the true heir to the 'proto-democracy' of archaic primitives, and this makes me suspicious of the whole procedure.

From *Nature* 27 March 1948.

Pharmacology

Substance P equals pain substance?

Leslie Iversen

Since von Euler and Gaddum first described substance P (SP) more than 60 years ago¹, much evidence has accumulated to suggest that this undecapeptide is released from the terminals of certain sensory nerves, as a chemical neurotransmitter or modulator². The sensory function of SP is thought, from more circumstantial evidence, to be related to the transmission of pain information into the central nervous system (CNS)^{2,3}. This idea is now supported by two studies reported by Cao *et al.*⁴ and De Felipe *et al.*⁵ on pages 390 and 394, respectively, of this issue. Both groups find that when the function of SP is genetically disrupted in mice, the animals show reduced responses to painful stimuli.

Cao *et al.*⁴ studied mice in which the preprotachykinin A gene was disrupted. This

gene encodes the precursor from which both SP and the closely related neuropeptide neurokinin A (NKA) are made. As expected, the mice lacked any detectable SP or NKA, although their development, ability to reproduce and behaviour were unaffected. The animals also showed normal thresholds for reactions to various painful stimuli, such as heat, mechanical pressure or chemical irritants. But when the intensity of the painful stimulus was increased, the knockout mice showed blunted responses, evidenced by increased latencies in the response to the stimulus. The importance of SP/NKA seemed to apply only to a certain 'window' of pain intensities — when the intensity of the pain stimulus was further increased, the responses of the knockout animals did not differ much from those of the wild-type mice.

De Felipe *et al.*⁵ studied a strain of mouse that lacked the neurokinin-1 (NK-1) receptor, with which SP interacts. These animals resembled those studied by Cao and colleagues in several respects: they did not show any changes in acute pain thresholds in mechanical, electrical or noxious heat tests, but their responses were blunted in tests that involved more intense noxious stimuli. When sensory nerves are subjected to an intense period of noxious stimulation, normal animals show a 'wind-up' phenomenon, whereby spinal reflexes are temporarily increased. This is thought to reflect the sensitization of CNS mechanisms by intense noxious stimulation. But such a wind-up was completely absent in the NK-1-receptor knockout mice.

As well as acting in the CNS, SP and other sensory neuropeptides can be released from the peripheral terminals of sensory nerve fibres in the skin, muscle and joints. This release is thought to be involved in 'neurogenic inflammation' — a local inflammatory response to certain types of injury or infection⁶. But the authors found that this process was impaired in both types of

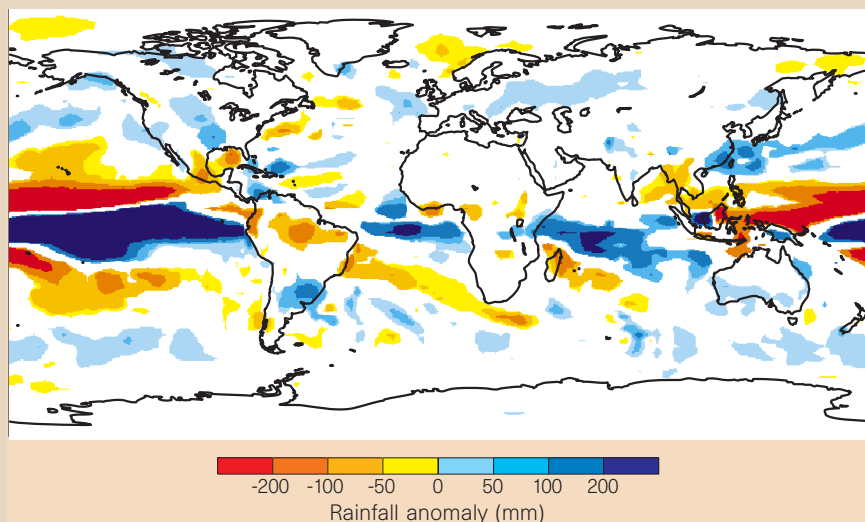
Climate forecasting

Outlook — probably wet in parts

Weather is too chaotic to be predicted more than ten days or so ahead. So why should we believe the global rainfall forecast, pictured here, made a few months in advance?

Well, this isn't really a weather forecast, in which the evolution of individual weather systems is predicted. It is a climate forecast, described elsewhere in this issue (Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S. & Balmaseda, M. A. *Nature* 392, 370–373; 1998), which uses a global model of the coupled oceanic and atmospheric circulations to calculate the statistical behaviour of large-scale weather patterns and the main physical influences on those statistics. The model calculates not only the mean (predictable) shift in climate for a season, but also estimates the random (unpredictable) component. Forecasts are therefore expressed as probabilities. For example, the model predicts that it is about 70% certain that southeast China will continue to experience the unusually wet weather it has been suffering of late.

The forecast was completed at the end of January 1998, and the map shows predicted total rainfall anomalies for this year's March–April–May season, relative to long-term mean rainfall for the same season; blue is wet, red is dry. Uncoloured areas represent a low-confidence forecast. As well as southeast China, the central United States and western central Eurasia



are predicted to be unseasonably rainy, whereas the forecast is that central Mexico and Indo-China will undergo shifts to an unusually dry season. Although some of these predictions are perhaps not so surprising in the wake of the present El Niño, Australia — which would normally suffer drought after an El Niño — seems to be in for a relatively wet time.

The high rainfall that contributed to last summer's severe and unexpected floods in central Europe was predicted using this approach. But a successful rainfall prediction is not necessarily a sign of a reliable forecasting system. On the one

hand, a forecast is only a probabilistic statement (so a good forecast might get the rainfall wrong); on the other, the model might forecast the correct rainfall for the wrong reasons. The European Centre for Medium-Range Weather Forecasts, where the map was produced, is investing substantial resources in addressing this 'verification' issue, largely by using past datasets to increase model confidence, and is now in a quasi-operational forecast phase. Go and look at the up-to-date seasonal forecasts (<http://www.ecmwf.int>), but don't bet your house on them.

Philip Newton

knockout animal. Moreover, the response to capsaicin was also absent or considerably reduced in the knockout mice. Normally, when this irritant is administered to the skin of the ear, it evokes a local oedema (accumulation of fluid) by provoking a release of sensory neuropeptides⁷.

Similar impairments of the neurogenic inflammatory response to capsaicin and the inflammatory response in the lung after administering an immune complex have already been reported in NK-1-receptor knockout mice by Bozic *et al.*⁸. On the other hand, the knockout animals studied by Cao *et al.* and De Felipe *et al.* responded normally to an inflammatory stimulus that did not involve a neurogenic component (the injection of complete Freund's adjuvant into a hind paw). Both the local inflammation and the delayed development of heightened pain sensitivity after these injections were unimpaired.

These new data come at an important time. Several pharmaceutical companies are developing drugs that act as potent antagonists of NK-1 receptors, although early results with such agents in animal tests of pain have often proved difficult to interpret. This is partly because the functional roles of SP and the NK-1 receptor are complex, and partly because some of the initial SP antagonists could not adequately penetrate the CNS or had other pharmacological activities that confused the results⁹. Nevertheless, the results with NK-1 antagonists mirror the present findings quite well — the drugs do not alter acute pain thresholds, but they do affect responses that involve more intense noxious stimulation. For example, they block the wind-up of spinal reflexes^{9–11}.

The results with the knockout mice and with the new SP antagonists show that SP is only one of the many complex mechanisms involved in pain perception. Mild and severe forms of pain involve other neurochemical mechanisms, among which L-glutamate is clearly important¹². De Felipe *et al.*⁵ also report that, in the NK-1-receptor knockout mice, the analgesic response to cold-water swim stress was considerably impaired. Moreover, these animals were much more aggressive than wild-type mice. The authors conclude "... that SP is important for orchestrating the response of the animal to major stressors such as pain, injury or invasion of territory". If they are right, then SP-receptor antagonists may have broad therapeutic applications in the treatment of a variety of stress-related illnesses, in addition to their potential as analgesics. □

Leslie Iversen is in the Department of Pharmacology, University of Oxford, Oxford OX1 3QT, UK.
e-mail: les.iversen@pharm.ox.ac.uk

2. Otsuka, M. & Yoshioka, K. *Physiol. Rev.* **73**, 229–308 (1993).
3. Hill, R. J. in *The Tachykinin Receptors* (ed. Buck, S. H.) 471–498 (Humana, Totowa, NJ, 1994).
4. Cao, Y. Q. *et al. Nature* **392**, 390–394 (1998).
5. De Felipe, C. *et al. Nature* **392**, 394–397 (1998).
6. Lembeck, F. & Gamse, R. *Ciba Found. Symp.* **91**, 35–54 (1982).
7. Caterina, M. J. *et al. Nature* **389**, 816–824 (1997).
8. Bozic, C. R., Lu, B., Höpken, U. E., Gerard, C. & Gerard, N. P. *Science* **273**, 1722–1725 (1996).

9. Longmore, J., Hill, R. G. & Hargreaves, R. J. *Can. J. Physiol. Pharmacol.* **75**, 612–621 (1997).
10. Laird, J. M. A., Hargreaves, R. J. & Hill, R. G. *Br. J. Pharmacol.* **109**, 713–718 (1993).
11. Ma, Q. P. & Woolf, C. J. *Eur. J. Pharmacol.* **322**, 165–171 (1997).
12. Dubner, R. & Basbaum, A. in *Textbook of Pain* (eds Wall, P. J. & Melzack, R.) 225–241 (Churchill Livingstone, London, 1994).

Palaeoclimatology

Icing the North Atlantic

Richard B. Alley

The North Atlantic region has experienced large, abrupt climate changes in the past¹, and similar changes occur in some models of the future². But there is disagreement as to whether these changes originate in the North Atlantic itself or are transmitted to the region from elsewhere. In their paper on page 373 of this issue³, McCabe and Clark tackle the question with new data on the extent and movement of the British Ice Sheet, which covered an area around the northern Irish Sea towards the end of the last ice age. From these data, they argue that the advance and retreat of one of the great ice sheets feeding the North Atlantic, the Laurentide, probably caused some of the big climate jumps of the past.

Two related oscillations are especially prominent in North Atlantic palaeoclimatic records — the lower-frequency Bond and the higher-frequency Dansgaard-Oeschger cycles (Fig. 1). Because signals of Dansgaard-Oeschger and, especially, Bond oscillations were recorded in geographically diverse areas extending well south of the Equator⁴, they should offer important clues in understanding and predicting climate changes.

In the Dansgaard-Oeschger cycle, rapid jumps (sometimes within less than ten years) of up to one-half of the glacial-interglacial amplitude in temperature and other climate variables occurred between climate states. Cold, dry and windy atmospheric conditions around the North Atlantic occurred when the ocean's surface waters were fresh and cold¹. Progressive cooling through several such Dansgaard-Oeschger cycles, followed by an especially abrupt warming, defines the Bond cycle. The coldest part of each Bond cycle in North Atlantic cores is marked by a so-called Heinrich layer rich in debris transported by icebergs¹ (Fig. 1).

These Heinrich layers thicken towards Hudson Bay, and the thicker parts were deposited anomalously rapidly. Ice-rafted debris in the thin marginal regions of Heinrich layers and between them has many sources from around the North Atlantic, but the thicker parts of Heinrich layers are dominated by material from Hudson Bay^{5–8}.

Many models for these oscillations

invoke processes taking place in ice sheets and the North Atlantic Ocean. Today, in an interglacial, cold and salty waters of the North Atlantic are dense enough to sink into the deep ocean, from where they flow south and are balanced by northwards surface flow from the tropics. Increased supply of fresh water in the north could dilute the northern surface waters sufficiently to prevent them sinking. Loss of the balancing, warm surface flow would cause high-latitude cooling and associated climate changes⁹, until salt build-up in the ocean or reduced freshwater supply reinitiated North Atlantic Deep Water formation. Forced or internally generated changes in freshwater delivery, possibly including warming-enhanced ice melt or precipitation, would create the observed

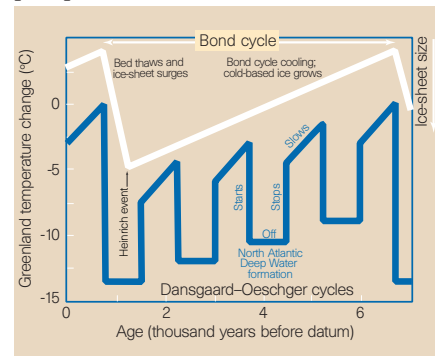


Figure 1 An idealized climate cycle in the North Atlantic region, assuming North-Atlantic-centred models for oscillating North Atlantic Deep Water formation¹⁰ and surging of the Laurentide Ice Sheet¹¹; time and temperature scales are approximate. The 1,500-year Dansgaard-Oeschger oscillations are related to weakening and shutdown of deep water formation as ice melt or other processes dilute the North Atlantic, and then resumption of deep water formation in response to the build-up of salt. The longer-term cooling of the Bond cycle is related to the downwind effects of the growing Laurentide Ice Sheet that covered Hudson Bay, and is terminated by a surge that thins the ice. Successive Bond cycles show warming or cooling in response to orbitally modulated ice-age cycles, and their coldest part is marked by a Heinrich layer. These are deposits in North Atlantic cores that consist of ice-rafted debris largely originating from the Hudson Bay area.

1. von Euler, U. S. & Gaddum, J. H. *J. Physiol. (Lond.)* **72**, 74–87 (1931).

Dansgaard-Oeschger signal¹⁰.

Growth of the Laurentide Ice Sheet in Hudson Bay through several Dansgaard-Oeschger oscillations would have trapped geothermal heat near the bed. Sufficient warming could have caused basal melting of the ice sheet and triggered an ice surge, flooding the North Atlantic with ice-rafted debris and melt water. The resulting Heinrich event would have initiated or strengthened a super-Dansgaard-Oeschger ocean cooling¹¹. The slow cooling and rapid warming of the Bond cycle may reflect the downwind effects of the slow growth, then rapid shrinkage of the ice sheet. Six or seven Heinrich layers over the most recent 100,000-year glacial cycle, most of which have strong connections with Hudson Bay⁵⁻⁸, may then record the repeated build-up and collapse of the great ice sheet there.

This is one point of view. However, the idea that changes at high latitudes can affect widespread regions extending across the Equator is unpopular with many workers who do not believe that the small, energy-starved polar 'tail' can wag the large, energy-rich tropical 'dog'. In this regard, Bond and Lotti¹² found that ice-rafted debris from diverse sources around the North Atlantic increased in sediment cores during each Dansgaard-Oeschger cooling, not just during the Heinrich events. They argued that a near-synchronous response of many ice masses is inconsistent with any stochastic model in which ice-sheet dynamics force climate change. Rather, they suggested that all of these ice masses were responding to some external signal, possibly from beyond the North Atlantic region, and that the Heinrich events are special only because something in the Laurentide Ice Sheet occasionally magnified the response.

McCabe and Clark³ now present data which suggest that the ice-rafted-debris signal of the youngest Heinrich event, called H1 and occurring about 17,000 years ago, is more consistent with a complex and varied response to a surge of the Laurentide Ice Sheet than with a common response of ice masses to an external forcing. With considerable (though not unequivocal) confidence, they show that the small British Ice Sheet advanced shortly after, and probably in response to, the cooling associated with Heinrich event H1. Cooling causes most small ice masses to advance because of reduced surface melt, often within years or decades¹³, so any cooling would be expected to increase the supply of ice-rafted debris to the North Atlantic from various sources. Further, an increase in ice-rafted debris may result from a glacial retreat if an ice-berg-calving margin suddenly collapses. Collapse could have occurred in response to a rise in sea level caused by a Laurentide surge, as McCabe and Clark argue happened with the ice sheets of the Barents Sea

and Fennoscandia. (These results may not exclude a more complex model, in which a Dansgaard-Oeschger-type cooling triggered a Heinrich event which triggered other responses^{1,14}.)

Correlations between Antarctic and Greenland ice cores, produced by Sowers and Bender¹⁵, show that warming of the Southern Hemisphere beyond the band of ice-age variability occurred while near-glacial cold conditions still prevailed in the north. This would be a major upset for the polar-forcing view of ice-age cycles, in which changes at high northern latitudes including the North Atlantic are transmitted to the south through the ocean and atmosphere¹⁶. But Sowers and Bender equally show that the first northern warming from the coldest conditions of the ice age led the first warming in the south. As McCabe and Clark emphasize, the anomalies then are the interruptions of northern warming, especially by Heinrich event H1 and again by the cold Younger Dryas interval of about 12,000 years ago (which has many characteristics of a Heinrich event and can be classed with them⁶), both plausibly linked to North Atlantic processes.

So, by building from an interesting new data set, McCabe and Clark refocus two of the important debates about deglaciation — the meaning of ice-rafted-debris deposits, and the interpretation of the north-south phasing of deglaciation. The tremendous rate of discovery on this topic almost guarantees that more exciting results will emerge soon, but McCabe and Clark have demonstrated that ice-sheet forcing of widespread climate change remains at least a viable model, and perhaps the preferred one. □

Richard B. Alley is in the Earth System Science Center and the Department of Geosciences, Pennsylvania State University, 204A Deike Building, University Park, Pennsylvania 16802, USA.

e-mail:ralley@essc.psu.edu

1. Bond, G. et al. *Nature* **365**, 143–147 (1993).
2. Stocker, T. F. & Schmittner, A. *Nature* **388**, 862–865 (1997).
3. McCabe, A. M. & Clark, P. U. *Nature* **392**, 373–377 (1998).
4. Lowell, T. V. et al. *Science* **269**, 1541–1549 (1995).
5. Grousset, F. E. et al. *Paleoceanography* **8**, 175–192 (1993).
6. Andrews, J. T., Erlenkeuser, H., Tedesco, K., Aksu, A. E. & Jull, A. J. T. *Quat. Res.* **41**, 26–34 (1994).
7. Gwiazda, R. H., Hemming, S. R. & Broecker, W. S. *Paleoceanography* **11**, 371–378 (1996).
8. McManus, J. F., Anderson, R. F., Broecker, W. S., Fleischer, M. Q. & Higgins, S. M. *Earth Planet. Sci. Lett.* (in the press).
9. Fawcett, P. J., Agustsdottir, A. M., Alley, R. B. & Shuman, C. A. *Paleoceanography* **12**, 23–38 (1997).
10. Broecker, W. S., Bond, G. & Klas, M. *Paleoceanography* **5**, 469–477 (1990).
11. MacAyeal, D. R. *Paleoceanography* **8**, 767–773 (1993).
12. Bond, G. & Lotti, R. *Science* **267**, 1005–1010 (1995).
13. Oerlemans, J. *Science* **264**, 243–245 (1994).
14. Alley, R. B., Anandakrishnan, S. & Cuffey, K. M. *Ant. J. US* (in the press).
15. Sowers, T. & Bender, M. *Science* **269**, 210–214 (1995).
16. Imbrie, J. et al. *Paleoceanography* **8**, 699–735 (1993).

Daedalus

Crookes flies high

The atmospheric mesosphere, between about 50 km and 100 km altitude, is too high for aircraft or balloons but too low for satellites. It borders the ionized layers which reflect radio waves, and generate aurorae. Daedalus now wants to explore it.

He has been inspired by the Crookes radiometer, that little windmill in a glass flask which spins in sunlight. Each vane is bright on one side but black on the other. The sunlight warms the blackened side of each vane, so that gas molecules hitting it rebound with extra momentum. This thermal transpiration exerts thrust on the vane. The effect peaks at around 5 pascals or so, just about the atmospheric pressure at 70 km altitude. So Daedalus is designing a big Crookes radiometer in the form of a helicopter rotor. Each angled aerofoil blade will be black on its upper surface, but reflective underneath. The blades will be angled so that sunlight shining down on them will drive them in that rotational sense that generates lift.

This huge frail craft, designed for tenuous high-altitude air, will have to be self-stabilizing. To keep its rotor horizontal, it will have a payload slung axially beneath its central hub. It will have enough lift to stabilize above its height of maximum efficiency. If it descends a little, it will then gain power and climb back up again. Oblique sunlight will not merely power it, but will urge it sideways, away from the Sun.

Thus, in the Northern Hemisphere, it will be pushed west and north during the morning, and east and north during the afternoon. At sunset it will lose lift; during the night it will spin lazily downwards like a sycamore seed. With luck it will still be within its operating altitude when the Sun rises next day; it will climb again, and start another leg of its zig-zag progress north. In summer, its odyssey will ultimately bring it into permanent polar sunlight, where it will orbit the pole steadily. When the changing seasons replace this cheerful buoyant radiance by winter darkness, it will fall to Earth and be lost.

Daedalus's altoradiometer will be a cheap, simple, long-lived research tool for polar aerospace. A balloon could just about lift it high enough for it to climb the rest of the way under its own power. Its solar-powered instrument package would then sample and transmit radiation fluxes, ion densities and magnetic data, for months on end. Astronomers could even observe stellar occultations through its fast-spinning blades.

David Jones

Obituary

Erik Jarvik (1907–98)

Palaeontologist renowned for his work on the “four-legged fish”

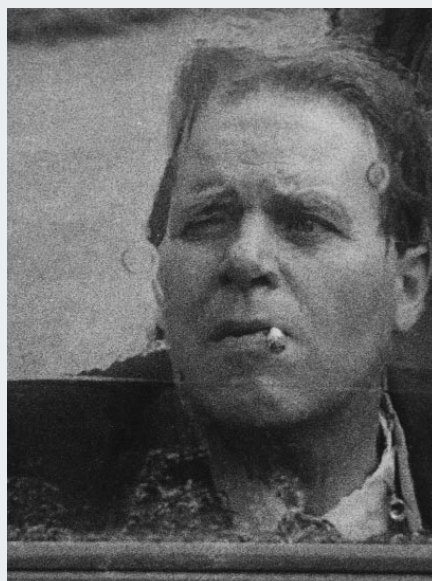
Erik Jarvik, who died on 11 January 1998, was one of the twentieth century’s leading researchers on early vertebrates and one of the main characters of the ‘Stockholm school’ of vertebrate palaeontology. His contribution to our knowledge of Palaeozoic fish and amphibians is unsurpassed, even though his views on their evolutionary relationships were largely regarded as heterodox by the wider scientific community.

Jarvik was born on a farm near Utby, a village in southwestern Sweden, and first studied natural sciences and geography at Uppsala University. He turned to vertebrate palaeontology after an expedition to East Greenland in 1932 with Gunnar Säve-Söderbergh, and spent all his career at the Swedish Museum of Natural History in Stockholm. By 1923, the museum’s palaeozoology department had become a world-famous centre for study of early vertebrates, thanks to Erik Stensiö, who had several foreign and Swedish disciples. Among them was Jarvik, who became a curator in the department in 1937 and was head of it from 1960 to 1972.

It has been said of Jarvik that he could be silent in seven languages. At meetings, his unusual theories on early vertebrate phylogeny, in particular the multiple origin of the four-legged land vertebrates (tetrapods), were often vividly opposed by colleagues. He rarely replied in public, but then, several years later, the response usually came in the form of a thoroughly argued paper, if not a huge monograph.

Jarvik liked to work in the large armchair in his Stockholm office, with a cup of Swedish boiled coffee at his side and a big cigar or a cigarette, painstakingly doing “his duty” — every day writing just half a page of some thoroughly elaborated text. This resulted in 51 papers, books and monographs, mostly on lobe-finned fish (a name he did not like, however) and tetrapods, but also on jawless vertebrates, basic vertebrate head structure, sharks and vertebrate origins. His book, *Basic Structure and Evolution of Vertebrates* (Academic, 1980, 1981), sums up his views and remains a highly valuable reference work.

Jarvik also took part in eight expeditions to Greenland and two to Spitsbergen between 1932 and 1969. In 1991, aged 84, he valiantly followed a field



party at Miguasha, in Canada, where his pet fossil fish, *Eusthenopteron foordi*, had been discovered in Devonian rocks.

Jarvik’s first important work was his doctoral thesis (1942) on the anatomy of the snout of the fossil lobe-finned (sarcopterygian) fish, which were already regarded as the possible ancestors of tetrapods. He noticed that, among these 370–400-million-year-old fossil fish, some (the porolepiforms) apparently shared the same snout structure with the urodeles (salamanders and newts), whereas others (the osteolepiforms) shared unique snout features with other tetrapods (frogs and amniotes). He thus suggested that tetrapods evolved twice, from two different fish groups, implying that limbs with digits appeared twice in vertebrate history. This polyphyletic theory of tetrapod origins was strongly opposed by most other palaeontologists. Nowadays, it is generally thought that tetrapods have a single origin; nonetheless, osteolepiform fish are certainly their close relatives.

For Jarvik, this work was decisive, because most of his later studies were aimed at supporting his theory. It led him to investigate the entire anatomy of fossil and Recent lobe-finned fish, and of the earliest known tetrapods that he collected in the 365-million-year-old rocks of East Greenland. His last work, published in 1996, was a monograph on the Devonian tetrapod *Ichthyostega*, one of the earliest known land vertebrates, which he liked to call the “four-legged fish” (a nickname introduced by a Danish cartoonist when *Ichthyostega* was first discovered in 1932 by Säve-Söderbergh). In the 1950s Jarvik’s descriptions of this animal raised great

interest among scientists and the public, because it was the first fossil evidence for an intermediate form between fish and tetrapods: it had limbs with digits but also a scaly fish tail with fin rays.

Jarvik’s theory of the polyphyletic origin of tetrapods rests on his way of assessing character homology (that is, shared characters inherited from a common ancestor) when reconstructing the relationships between living and fossil groups. For Jarvik, two widely different characters in distant groups could be regarded as homologous as long as they belonged to the same segment of the ideal vertebrate head and shared the same embryonic derivation. Conversely, minor anatomical differences were sometimes regarded by him as evidence for widely different ‘bauplans’ (basic structures). Being either too broad or too narrow, this concept of homology often could not be used for phylogeny reconstruction. It is not surprising that such views, expressed at the dawn of evolutionary systematics in the 1940s or 1950s, were regarded as heretical.

In the late 1960s, with the advent of Willi Hennig’s phylogenetic systematics (cladistics, which dethroned evolutionary systematics in the 1970s), Jarvik readily accepted Hennig’s principles of phylogeny reconstruction and classification, because they embodied several ideas that had long been held by the Stockholm school, such as the rejection of primitive characters in phylogeny reconstruction, the unknowability of actual ancestors, the unreliability of the stratigraphical record, and the importance of embryonic development in assessing the primitive or advanced states of characters. Why Jarvik soon after rejected this major renewal in comparative biology remains obscure. It may be because he considered it was nothing new to him, or because he rejected the principle of parsimony (the cornerstone of cladistics), which threatened his polyphyletic theory of tetrapod origins.

Nevertheless, whether Jarvik was right or wrong about certain questions of vertebrate phylogeny is no longer important. What we now see as fundamental in his work is his constant care for the interpretation of fossils in the framework of the anatomy and embryology of their living relatives, based on observations and descriptions of outstanding quality.

Philippe Janvier

Philippe Janvier is in the Laboratoire de Paléontologie, Muséum National d’Histoire Naturelle, URA 12 CNRS, 8 Rue Buffon, 75005 Paris, France.
e-mail: janvier@cimrs1.mnhn.fr

Microbial genomes opened up

Russell F. Doolittle

The complete genome sequences of a dozen bacteria, and a yeast, provide a wealth of information for tracing evolutionary networks. Here is a brief guide to what has and can be learned.

It is less than three years ago that the first complete sequence of a bacterial genome was reported¹. Now we have a dozen¹⁻¹², and that of a eukaryote¹³ (the budding yeast, *Saccharomyces cerevisiae*), with a certainty of many more to follow (Table 1).

One of the most exciting aspects of this cavalcade of genomes is the immediate public access to the data. The microbial genome database Web site at The Institute for Genome Research (TIGR, sited in Maryland) currently lists almost 50 projects under way around the world. Half of these are being conducted by TIGR, either alone or in collaboration with others. Assorted universities around the world have staked out particularly interesting organisms; some of the projects are being undertaken as consortia, international or national (for example, a Brazilian consortium is sequencing the plant pathogen *Xylella fastidiosa*). The Web site posting for the Sanger Centre, Cambridge, UK, currently shows five microbial genomes in progress, and — reminiscent of aircraft departure monitors in large airports — the entry for *Mycobacterium tuberculosis* continuously flashes 'Completed'.

But what is so sensational about having a 'complete sequence' instead of only the sequences of the 'interesting' genes, many of which have been reported long ago? For one thing, what is absent is sometimes as interesting as what is present; until an entire genome is in hand one can't be certain that some other functionally or otherwise related gene isn't lurking behind the scenes. Beyond that, there are some fundamental questions that these studies might be expected to answer, none of which has yet been answered definitively, mostly because identifying all of the genes on the basis of sequence alone has proved more difficult than expected. Additionally, there is increasing confusion about the relationship of the major divisions of life.

The tree of life

Until the late 1970s, living organisms were divided into two major groups: the prokaryotes (no nucleus) and the eukaryotes (defined nucleus). The advent of RNA sequencing led to a three-kingdom world¹⁴, the prokaryotes being divided into two groups: the true bacteria (Eubacteria, also known as Bacteria) and a motley group composed of organisms from diverse and extreme

habitats (the Archaeobacteria, a.k.a. Archaea)¹⁵. Three of the first dozen bacterial genomes to be completed belong to the Archaea, allowing for a thorough comparison of the two groups.

Although a phylogenetic tree based on ribosomal RNA from the 13 organisms with fully sequenced genomes nicely bears out the three-kingdom world (Fig. 1, overleaf), not all of their gene products are so accommodating. Remarkably, in spite of many expectations to the contrary, the vast majority of gene products from the Archaea most resemble counterparts among the Eubacteria and not eukaryotes^{8,10}. Even so, that a significant minority of archaeal proteins, and especially those having to do with the transcription and translation of other genes, more closely resemble eukaryotic forms, reinforced notions that eukaryotes were the result of a chimaeric merging of a eubacterium and an archaeobacterium¹⁶⁻²⁰, the most recent report of which proposes a symbiosis between a eubacterium that consumed organic matter and excreted H₂ and CO₂, and a methane-generating archaeobacterium that used these by-products for food²¹. Whatever the case, it is almost certain that the last common ancestor of Bacteria and Archaea was also the last common ancestor of all extant life, although the origin of the Eukarya is far from settled.

The Human Genome Project

The microbial genome projects, like those for a number of multicellular creatures, have roots that reach back to the Human Genome

Project. Only a decade ago the biological community was debating the feasibility of sequencing the human genome, a bothersome question being how one would recognize genes on the basis of DNA sequence alone. Setting aside the non-coding sequences dubbed introns, what is it that would allow the identification of 50,000–100,000 genes in the human genome when, at the time, fewer than a thousand gene products for humans had been sequenced?

In the end, sceptics were persuaded by the realization that all living things are the result of an enormous genetic expansion. New genes (and the proteins they encode) come from old genes by way of duplication and modification, and, for the most part, these modifications occur at a slow enough pace that related genes are easily recognized on the basis of sequence alone.

Briefly put, the dogma was that all creatures on Earth are descended from a common ancestor and share a fundamental set of genes, the translated sequences of many of which are easily recognized even between prokaryotes and eukaryotes. For example, the amino-acid sequences of the enzyme triose phosphate isomerase from humans and the bacterium *E. coli* were already known to be 45% identical. As a result, for many proteins one doesn't need to have the human sequence already in hand; it will be easily identified on the basis of sequences that have already been found in other creatures.

There was also concern that the human genome has many more genes than many of the organisms that molecular biologists had traditionally concentrated on, such as *E. coli* and yeast, or even the nematode worm *Caenorhabditis elegans* and the fruitfly *Drosophila melanogaster*. The fear was that vast numbers of genes would be found that had no counterparts in the data banks. These concerns gradually diminished as appreciation of gene duplications grew, and 'gene families' became a part of the molecular biologist's lexicon.

Table 1 Some completely sequenced genomes

Organism	Size (Mb)	ORFs	Year reported	Web site
<i>Mycoplasma genitalium</i>	0.58	470	1995	www.tigr.org/tdb/mdb/mgdb/mgdb.html
<i>Mycoplasma pneumoniae</i>	0.82	679	1996	www.zmbh.uni-heidelberg.de/M_pneumoniae/
<i>Borrelia burgdorferi</i>	0.91	843	1997	www.tigr.org/tdb/mdb/bbdb/bbdb.html
(<i>Borrelia</i> plasmids)	(0.32)	(430)	1997	www.tigr.org/tdb/mdb/bbdb/bbdb.html
<i>Aquifex aeolicus</i>	1.55	1,512	1998	(available through GenBank)
<i>Helicobacter pylori</i>	1.67	1,590	1997	www.tigr.org/tdb/mdb/hpdb/hpdb.html
<i>Methanococcus jannaschii</i>	1.67	1,738	1996	www.tigr.org/tdb/mdb/mjdb/mjdb.html
<i>M. thermoautotrophicum</i>	1.75	1,855	1997	www.genomecorp.com/htdocs/sequences/
<i>Haemophilus influenzae</i>	1.83	1,743	1995	www.tigr.org/tdb/mdb/hidb/hidb.html
<i>Archaeoglobus fulgidus</i>	2.18	2,436	1997	www.tigr.org/tdb/mdb/afdb/afdb.html
<i>Synechocystis</i> sp.	3.57	3,168	1996	www.kazusa.or.jp/cyano/cyano.html
<i>Bacillus subtilis</i>	4.21	4,100	1997	www.pasteur.fr/Bio/Subtilist.html
<i>Escherichia coli</i>	4.67	4,288	1997	www.genetics.wisc.edu:80/
<i>Yeast (S. cerevisiae)</i>	12.07	5,885	1996	www.mips.biochem.mpg.de/mips/yeast

The status of many other microbial genome projects under way can be checked at www.tigr.org/tdb/mdb/mdb.html or at www.sanger.ac.uk/ and by clicking on 'microbes'.

A boost from ESTs

Another concern about identifying human genes on the basis of sequence alone had to do with the fragmentation of genes in 'higher' eukaryotes, whereby coding information is interrupted by the non-coding intron sequences. Some critics felt that it was unwise to sequence genomic DNA, only about 5% of which is coding and actually expressed as proteins. Why not stick to complementary DNA, they argued, where identifications will be easier? (cDNA is the copy of messenger RNA without introns, and directly encodes the amino-acid sequences of proteins.)

The point was forcefully made in 1991 when a group from TIGR reported the initial results of a massive sequencing effort on a large random cDNA library from human brain cells²². Instead of fully characterizing every cDNA clone, they simply used an array of automatic DNA sequencers to examine individual but randomly selected clones, subjecting each to a single sequencer run, usually averaging about 450 bases, the equivalent of 150 amino acids. Because cDNA clones are prepared from messenger RNA, and thereby represent genes that are being expressed, these sequences were called expressed sequence tags, or ESTs. The TIGR group searched the databases with these partial sequences, finding related sequences for tens of thousands of them, and thereby allowing good estimates of the expected gene count for humans²³.

Indeed, it was the power of having several dozen machines accurately churning out a combined equivalent of 500 kilobases a day that set the stage for TIGR to undertake the sequencing of a complete bacterial genome. This brute force approach was coupled with solid Poisson-based expectations of how many sequences of 400–500 base pairs in length would be needed for complete

coverage of a bacterial genome by shotgun sequencing. Thus, the first complete bacterial genome to be reported, the 1.83-megabase sequence of *H. influenzae*¹, was determined by randomly sequencing fewer than 20,000 clones, a sixfold redundancy, on the average, for every base, and one that had the advantage of simultaneously providing an accuracy and reliability check (if one of the six gave a different base than the others, it could be checked).

Once a genome's sequence is determined, the analysis begins in earnest. Although there is interesting information at every turn, the main priority is to identify the genes. This process is easier in bacteria than in eukaryotes because their genes lack those nuisancesome introns. Simple computer programs translate the DNA sequence directly into putative protein sequences. Long runs of coding triplets without stop codons are called 'open reading frames', or ORFs, and presumably correspond to the genes for proteins.

The ORF sequences are then searched against large databases of all known sequences from all kinds of organisms to see if they resemble any known protein. Surprisingly, a substantial number of the ORFs uncovered in these first complete genomes have not looked like any reported protein, which makes them URFs (unidentified reading frames), as well as ORFs (Fig. 2).

Recognizing genes

As a rule of thumb, any two protein sequences that are more than 25% identical are likely to be homologous — that is, descended from a common ancestor. If the resemblance is less, one must be more cautious; the region between 15% and 25% identity is often referred to as the 'twilight zone'. But proteins whose sequences are even less similar can be shown to be related, either

by pattern analysis or by comparing their three-dimensional structures.

The large fraction of unidentified genes is a major stumbling block to answering basic questions about evolution. Some of these URFs may encode genes common to all organisms, but their sequences alone may not be sufficient to make their identity known; or they may be the genes that impart uniqueness to the organism. And not all of these URFs, especially the shortest among them, necessarily get translated into functional proteins.

Is a particular ORF unidentifiable simply because it encodes a very fast-changing protein, the sequence of which has been blurred by numerous amino-acid substitutions? Or does it represent a gene whose function has not yet been found by biochemists and molecular biologists? If it is the latter, then it appears that almost half of cell biology and biochemistry has not yet been encountered, in spite of a century of exploration.

With so many complete genome sequences now on hand, the two possibilities ought to be distinguishable by direct genome comparison. Thus, if the same URF is recognizable in the genomes of distantly related bacteria, it is not likely to be a case of non-identification due to rapid change. On the other hand, if an URF is unique to an organism, it could be that either it has changed so much that it can't be recognized, or it might really be a 'uniqueness gene'. In this case, the matter can be sorted out if sequences are available from two or more closely related organisms where rates of change can be compared with other genes.

Little and large microbes

The prokaryotic genomes determined so far range in size from about 0.6 to 4.7 megabases (Table 1). Inherent in this distribution is the evolutionary potential to expand or shrink. Expansion is made possible by gene duplications, and shrinkage by deletions. It is of considerable interest that different clusters of duplicated genes are showing up in the various genomes. For example, in *B. subtilis* more than a quarter of the genes show evidence of recent duplication; a hundred of them are present in paralogous sets of five (a paralogous gene or protein is one that has homology to another as a result of gene duplication). In *E. coli*, there is a family of 80 paralogous transport proteins⁷. Tallying up all the recent duplications in these genomes may result in some general rules about frequency of occurrence.

Similarly, the absence of many genes in smaller genomes can provide hints about the ease of deletion and removal of genes that are no longer needed. Among the bacteria that have had their genomes completely sequenced, for example, are two closely related organisms of the genus *Mycoplasma*, whose members have the smallest cellular genomes

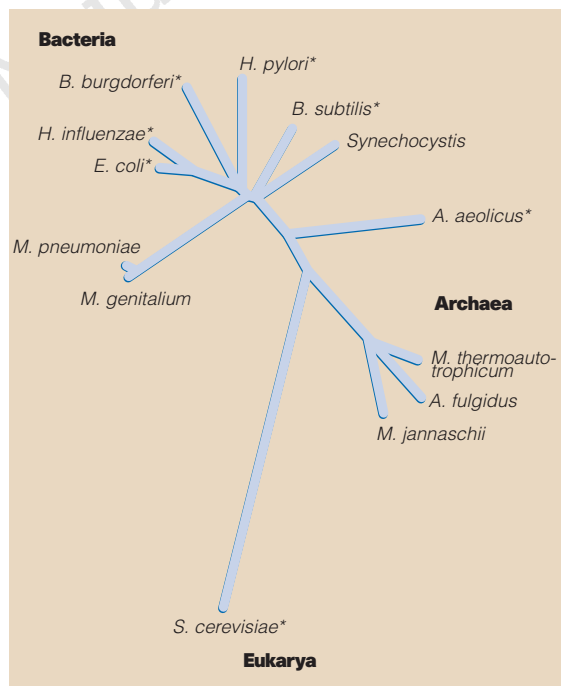


Figure 1 The three-kingdom world — a phylogenetic tree based on sequences of ribosomal RNA from the 12 prokaryotes and one eukaryote described in Table 1. Organisms marked with an asterisk have the actin-related protein ftsA, or, in the case of yeast, actin itself. Note that ftsA is not found in the Archaea. The significance of this is to do with the origin of complex eukaryotes from prokaryotes, and the mystery of how they acquired cytoskeletal proteins such as actin.

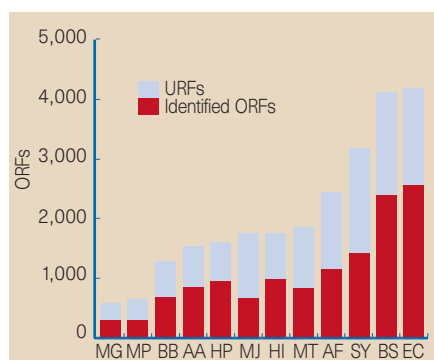


Figure 2 Open and unidentified reading frames (ORFs and URFs) in 12 completely sequenced bacterial genomes. MG, *Mycoplasma genitalium*; MP, *Mycoplasma pneumoniae*; BB, *Borrelia burgdorferi*; AA, *Aquifex aeolicus*; HP, *Helicobacter pylori*; MJ, *Methanococcus jannaschii*; HI, *Haemophilus influenzae*; MT, *Methanococcus thermoautotrophicum*; AF, *Archaeoglobus fulgidus*; SY, *Synechocystis* sp.; BS, *Bacillus subtilis*; EC, *Escherichia coli*.

known; this is a reflection of their parasitic existence, because many of their needs are provided by their animal hosts. *Mycoplasma genitalium* has only 470 ORFs and *M. pneumoniae* 679. All 470 ORFs from the smaller bacterium are found in the larger relative. Their protein sequences are about 67% identical, on average, a reflection of their close relationship²⁴.

The two mycoplasmas also share numerous URFs that are not recognized in any of the other fully sequenced genomes. One of these is a 'hypothetical protein' that is more than 90% identical in the two organisms; clearly, this isn't a case of a fast-changing protein. So, where did it come from? Similarly, there are URFs found only in the larger bacterium (*M. pneumoniae*) that have been observed, and remain unidentified, in other distantly related bacteria such as *E. coli* and *B. subtilis*. Again, it isn't a matter of rapid change that is keeping their identities secret.

As for what makes mycoplasma unique, 350 of the 470 ORFs found in *M. genitalium* have corresponding relatives in *B. subtilis*⁹, leaving only about a hundred candidates for 'uniqueness genes'. Interestingly, the next smallest of the fully sequenced genomes, that of *B. burgdorferi*, is also an animal parasite that lacks many of the same enzymes and metabolic pathways that are absent in the mycoplasmas, even though it is distantly related (Fig. 1).

For example, both of these organisms lack biosynthetic capabilities for amino acids, fatty acids, nucleotides and enzyme co-factors, and none of the enzymes of the tricarboxylic acid cycle is found in either of them, leading to the notion of 'convergent evolutionary gene loss'¹¹. Viewed in retrospect, this isn't so surprising. If an organism is provided with an abundant supply of nutrients, and

has the wherewithal to absorb them, then it seems predictable that the machinery responsible for their synthesis would decay. It is an old rule of evolutionary biology and natural selection: 'Use it or lose it'.

Minimal gene content

As soon as two bacterial genomes became available, estimates were made of 'minimal gene content'. Comparison of the genomes of *H. influenzae* and *M. genitalium* led to the claim that a set of 256 genes is "close to the minimal necessary and sufficient to sustain the existence of a modern-type cell"²⁵. That both of these organisms are parasitic on hosts which provide much of their sustenance confounds the significance of such a conclusion, even though both bacteria can be cultured in the absence of their hosts. Nevertheless, this is a provocative suggestion that can be tested experimentally. It would be notable indeed if molecular biologists deleted all but the 256 genes from *M. genitalium* and found that the organism remained viable.

The last common ancestor

I worry that many readers will not appreciate the distinction between the alleged minimal gene content and the quite different matters of the gene content of the common ancestor, on the one hand, or the number of genes in the earliest cells, on the other. *Haemophilus influenzae* and *M. genitalium* are quite distantly related, the former being a Gram-negative bacterium and the latter Gram-positive (this is one of the fundamental divides in the bacterial world). At the most recent, they last shared a common ancestor two billion years ago²⁶.

Clearly, the list of 256 genes they share in no way reflects the common ancestor, as can be shown by considering all other gene products shared by other similarly related bacteria. For example, *B. subtilis* and *E. coli* — also Gram-positive and Gram-negative, respectively — share large numbers of genes not on the minimal list of 256, and the last common ancestor of *H. influenzae* and *M. genitalium* must also have had these genes. Although a comprehensive count of all such occurrences has yet to be made, the last common ancestor of Gram-positive and Gram-negative bacteria probably had a thousand or more genes.

Such an analysis also differs from looking at what genes are shared by all the known genomes. In the first nine bacterial genomes compared rigorously, only 34 genes are certain to be common to all of them²⁷. The difference between the numbers shared by any pair and those shared by the whole group reflects differential loss along different lineages. It is the total number shared by all members of the archaeobacteria with all eubacteria that needs to be used for estimating the gene composition of the last common ancestor of all life, barring one other confounding circumstance: the awkward matter of horizontal gene transfers.

Horizontal gene transfers

The underlying rationale for phylogenetic analysis with macromolecular sequences depends on the common ancestor having a defined sequence of some significant length, and that the sequence changes gradually and differently along descendant limbs of a divergence as the result of single base substitutions and small insertions and deletions. As a result, the corresponding sequences of descendants differ according to their relatedness: the more different the sequences of two extant organisms, the more distant their evolutionary relationship. Still, occasionally phylogenies are generated that are completely out of keeping with other phylogenies, and often as a last resort 'horizontal gene transfer' is invoked.

The evidence for such transfer between distantly related bacteria is strong, however. It falls into two realms. First, there are the sequence comparisons themselves. The sequence of the enzyme dihydrolipoamide dehydrogenase from the archaeobacterium *Halobacterium halobium* is much more like those of Gram-positive eubacteria (50% identical) than it is like any sequence from the three fully sequenced Archaea (25% identical); the simplest explanation is a gene transfer.

Another example is the adenyl sulphate reductase from the archaeon, *A. fulgidus*, the sequence of which is very similar to an orthologue found in sulphite-reducing proteobacteria, but absent in all other Archaea. (Orthologous proteins in different organisms are related by common ancestry, and have the same function. The differences between them result from speciation.) Moreover, a set of three sulphite reductase genes makes up an operon, or gene cluster, in *A. fulgidus* that exactly parallels those found in the sulphate-reducing proteobacteria¹⁰. All of these observations imply that horizontal transfer has taken place.

There are also ancillary considerations that support the existence of horizontal transfers. The guanine+cytosine content of a gene may differ considerably from its surroundings, or the tracks of insertion sequences may be present, or there may be evidence of prophages — forms of bacterial virus that have been integrated into the host bacterium's DNA. In *H. pylori*, for example, the 'pathogenicity island', which is one of the mainstays of the organism's virulence, is delineated by 31-base-pair direct repeats⁶, and in *B. subtilis* there are ten prophages or their remnants, evidence of past horizontal transfers⁹. Also, the scattered distribution of intervening amino acid sequences that are spliced out of immature proteins (inteins) in Archaea is testimony to horizontal transfers; *M. jannaschii* has 18 inteins, *M. thermoautotrophicum* only one, and *A. fulgidus* none at all¹⁰.

Shuffling genes

The order in which the genes occur on a bacterial chromosome is not entirely random.

Indeed, genes that are involved in a particular function often occur next to one another and are coordinately regulated as 'operons' — clusters of genes associated with the same physiological function and transcribed on a single messenger RNA. Operons are common in the genomic sequences of both Bacteria and Archaea, but they may occur in widely different locations in different organisms²⁸. The implication is that, although the bacterial genome is experiencing constant reshuffling, a consequence of the persistent breaking and rejoining of DNA, there is a natural advantage to keeping some genes near each other.

There are, however, differing interpretations of what that advantage might be. The classical view has been that it is primarily a matter of regulating gene expression²⁹. A couple of years ago, that view was challenged with the publication of the hypothesis of the 'selfish operon'³⁰. This view is based on the notion that horizontally transferred genes will fare better if they are accompanied by other genes with whose products they interact. The contention is that coordinate gene regulation need not require genes to be in proximity to one another.

On the other hand, an analysis of those operons that have been most preserved in the fully sequenced genomes has shown that they are predominantly of the kind regulated by mechanisms at the RNA level. It has been suggested that these operons reflect a kind of gene regulation that may pre-date the invention of DNA and that dates back to an ancestor with an RNA genome²⁸.

Missing genes

In almost all of the fully sequenced genomes there are cases of 'missing' ORFs — that is, biochemical activities have been observed, but the genes responsible have not been found. By far the most interesting case has to do with a missing lysyl transfer RNA synthetase in archaeobacteria. This is the enzyme responsible for incorporating the amino acid lysine into proteins. Thus, when the complete genome of the archaeobacterium *M. jannaschii* was determined, it was found that four of the 20 amino-acyl-tRNA synthetases could not be identified. There was good reason to think that the enzymes for glutamine and asparagine would not be present, as previous work had shown that in archaeobacteria and many eubacteria these amino acids are incorporated as transamidated derivatives of glutamate and aspartate, respectively³¹. That the enzyme for cysteine was not found was puzzling, but the chemical relatedness of cysteine and serine allowed the possibility that a trans-sulphuration reaction might occur on a charged serine-tRNA. That the lysine enzyme could not be identified was more baffling, however, because lysine cannot be readily generated from any other amino acid.

Late last year, the tRNA synthetase for lysine was found biochemically, cloned and

identified in the *M. jannaschii* genomic sequence³². The first surprise was that it looks not at all like other lysine-tRNA synthetases, and belongs, so far as can be told, to a different class. The initial reaction was that somehow this enzyme had changed very rapidly in evolution, to the extent that it was no longer recognizable. But, even before a careful comparison of evolutionary rates could be made, a second surprise occurred: computer searching of the newly determined *B. burgdorferi* genome revealed the same kind of lysine-tRNA synthetase³³! The lysine-tRNA synthetase sequences from the distantly related *Methanococcus* and *Borrelia* (Fig. 1) are about 30% identical, which is just about the same as for their other tRNA synthetases.

Rapid rate of change is not the explanation. Instead, what has most likely occurred is a gene displacement, several examples of which have been observed in the past. For example, bacterial ornithine decarboxylases are much more like bacterial lysine decarboxylases than eukaryote ornithine decarboxylases, and bacterial tyrosine transaminases are more like bacterial aspartate transaminases than eukaryote tyrosine transaminases³⁴. In each case it appears that gene duplication has led to a paralogue displacing an orthologue.

Sometimes the displacement is not an obvious paralogue, the new gene product having virtually no resemblance at all to the displaced agent. To cover both kinds of change, Koonin *et al.*³⁵ introduced the somewhat redundant term 'non-orthologous displacement' (if I quibble with this terminology, it's because I can't envisage an orthologous displacement). Nomenclature aside, these authors have systematically identified a number of such displacements, including the possible displacement of a missing nucleoside phosphokinase by another kind of kinase³⁵. More recently, several proteins have been identified in the *B. subtilis* genome that have the same function as completely different proteins in *E. coli*⁴⁰.

Eukaryote origins

Of all the missing genes, the ones I find most perplexing are those of the eukaryotic cytoskeleton — the framework of protein filaments that give a cell its shape and ability to move³⁶. Where are the precursors of tubulin and actin and other cytoskeletal proteins? So although much has been made recently about the resemblance of the prokaryotic cell-division protein ftsZ to tubulin from eukaryotes³⁷, phylogenetic trees based on these proteins cluster the Archaea and Bacteria firmly together and far from the tubulin-based position of eukaryotes. It is even more disappointing that another prokaryotic cell-division protein thought to be related to actin³⁸, ftsA, isn't in any of the archaeal genomes (Fig. 1). The absence of sequences closely related to the slowly changing proteins of the eukaryotic

cytoskeleton remains unsettling, and the origin of the cytoskeleton cannot easily be accounted for by simple chimaeric mergers of a eubacterium and an archaeobacterium.

Rather, as the evolutionary trees based on ribosomal RNA have always implied, there must have been a third party involved in the origin of complex eukaryotes. Isn't it possible, for example, that the suggested merger of a eubacterium that excreted H₂ and CO₂ with an archaeobacterium that consumed those by-products²¹ actually took place within a third cell that already had a cytoskeleton? Were the forebears of that ancient third lineage, unlike the relatives of the swallowed Bacteria and Archaea, completely overrun by the trimeric entity?

There may be important clues in the genomes of Eukarya more primitive than yeast. The TIGR Web site reveals that eight different groups are at work on chromosomes from the protist *Plasmodium falciparum*. Quite apart from the medical implications of having the sequence of the organism that causes malaria, perhaps some questions about the origin of the eukaryotic cell will be answered as well. □

Russell F. Doolittle is in the Center for Molecular Genetics, University of California, San Diego, La Jolla, California 92093-0634, USA.
e-mail: rdoolittle@ucsd.edu

1. Fleischmann, R. D. *et al.* *Science* **269**, 496–512 (1995).
2. Fraser, C. M. *et al.* *Science* **270**, 397–403 (1995).
3. Bult, C. J. *et al.* *Science* **273**, 1058–1073 (1996).
4. Kanehisa, T. *et al.* *DNA Res.* **3**, 109–136 (1996).
5. Himmelreich, R. *et al.* *Nucleic Acids Res.* **24**, 4420–4449 (1996).
6. Tomb, J.-F. *et al.* *Nature* **388**, 539–547 (1997).
7. Blattner, F. R. *et al.* *Science* **277**, 1453–1462 (1997).
8. Smith, D. R. *et al.* *J. Bacteriol.* **179**, 7135–7155 (1997).
9. Kunst, F. *et al.* *Nature* **390**, 249–256 (1997).
10. Klenk, H.-P. *et al.* *Nature* **390**, 364–370 (1997).
11. Fraser, C. M. *et al.* *Nature* **390**, 580–586 (1997).
12. Deckert, G. *et al.* *Nature* **392**, 353–358 (1998).
13. Goffeau, A. *et al.* *Science* **274**, 546–567 (1996).
14. Woese, C. R. & Fox, G. E. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090 (1977).
15. Woese, C. R., Kandler, O. & Wheelis, M. L. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
16. Hartman, H. *Sci. Technol.* **7**, 77–81 (1984).
17. Sogin, M. L. *Curr. Opin. Genet. Dev.* **1**, 457–463 (1991).
18. Puhler, G. *et al.* *Proc. Natl Acad. Sci. USA* **86**, 4569–4573 (1989).
19. Gupta, R. S. & Golding, G. B. *J. Mol. Evol.* **37**, 573–582 (1993).
20. Koonin, E. V. *et al.* *Molec. Microbiol.* **25**, 619–637 (1997).
21. Martin, W. & Müller, M. *Nature* **392**, 37–41 (1998).
22. Adams, M. D. *et al.* *Science* **252**, 1651–1656 (1991).
23. Adams, M. D. *et al.* *Nature* **355**, 632–634 (1992).
24. Himmelreich, R. *et al.* *Nucleic Acids Res.* **25**, 701–712 (1997).
25. Mushegian, R. & Koonin, E. V. *Proc. Natl Acad. Sci. USA* **93**, 10268–10273 (1996).
26. Feng, D.-F., Cho, G. & Doolittle, R. F. *Proc. Natl Acad. Sci. USA* **94**, 13028–13033 (1997).
27. Huynen, M. A. & Bork, P. *Proc. Natl Acad. Sci. USA* (in the press).
28. Siefert, J. L. *et al.* *J. Mol. Evol.* **45**, 467–472 (1997).
29. Jacob, F. & Monod, J. *J. Mol. Biol.* **3**, 318–356 (1961).
30. Lawrence, J. G. & Roth, J. R. *Genetics* **143**, 1843–1860 (1996).
31. Ibba, M., Curnow, A. W. & Soll, D. *Trends Biochem. Sci.* **22**, 39–42 (1997).
32. Ibba, M. *et al.* *Science* **278**, 1119–1122 (1997).
33. Ibba, M. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 14383–14388 (1997).
34. Doolittle, R. F. *et al.* *Science* **271**, 470–477 (1996).
35. Koonin, E. V., Mushegian, A. R. & Bork, P. *Trends Genet.* **12**, 334–336 (1996).
36. Doolittle, R. F. *Phil. Trans. R. Soc. Lond. B* **349**, 235–240 (1995).
37. Lowe, J. & Amos, L. A. *Nature* **391**, 203–206 (1998).
38. Bork, P., Sander, C. & Valencia, A. *Proc. Natl Acad. Sci. USA* **89**, 7290–7294 (1992).

Turner's Trinity

Turner is famous as a master of the use of primary colours to depict light in painting. What is less well known is that he was also a student of the science behind that art.

Martin Kemp

When J. M. W. Turner was asked to explain his pair of paintings exhibited at the Royal Academy in 1843, *Shade and Darkness — the Evening of the Deluge* and *Light and Colour (Goethe's Theory) — the Morning After the Deluge*, he tersely responded, "Red, blue and yellow". This triad comprised painters' traditional 'primaries', from which (in theory) all other colours could be mixed.

Looking at the swirling chaos of colour in Turner's vision of the Deluge's aftermath, it is difficult to believe that it is founded on scientific ideas, yet he cites "Goethe's Theory" — the anti-Newtonian doctrines expounded in Goethe's *Farbenlehre* in 1810.

Goethe, although on the eccentric fringe of science, was far from alone at that time in reviving the three-colour theory in the face of Newton's apparently triumphant doctrine of seven unrefrangible hues.

For example, Sir David Brewster, the Scottish scientist and biographer of Newton, polemically advocated what he called the "Trinity" of colours. Relying upon observations of the decomposition of light by absorption through coloured filters, Brewster argued that white light was produced by three overlapping spectrums of red, yellow and blue.

Although Turner, who met Brewster, realized that actual mixtures of "aerial" colours (additive mixing of coloured lights) and "material" colours (subtractive mixtures of pigments) produced quite different results, he welcomed the idea that there was a direct correspondence of hue between the primaries for lights and pigments.

Given Turner's status as the supreme exponent of Romanticism in British art, his avid interest in scientific theory comes as something of a surprise. He devoted long



Colour Circle No. 1: diagram of the "aerial colours" from Turner's lectures (Tate Gallery, London).



J. M. W. Turner's *Light and Colour (Goethe's Theory) — The Morning After the Deluge*, 1843 (Tate Gallery, London).

hours to the preparations for his lectures as professor of perspective at the academy, where he introduced his audience to Euclidean geometry and the complex issues surrounding the strengths and fallacies of conventional perspective. He was conscious that each portion of space depicted in perspective was a fragment of an infinite whole, and that our eyes embrace an orbital sweep of concave space, dynamically evoked in the Goethe canvases.

Scientific insight, for Turner, was never a matter of 'restrictive rule', but provided access to the awesome powers of nature. He was fascinated by the magnetizing properties of light, and by the action of sunlight on silver salts as exploited to such effect by the pioneers of photography.

Turner did not swallow Goethe's theory whole. "Poor Dame Nature" he wrote in one of his annotations to his copy of the 1840 translation of *Farbenlehre*, when he felt that

Goethe was downgrading nature's supreme powers. But he was attracted to the way the German set his two "polar" colours, yellow and blue, in concert with paired forces of nature: plus and minus; action and negotiation; force and weakness; warmth and coldness; proximity and distance; repulsion and attraction; acids and alkalis, and so on.

In Turner's pair of paintings, the dark, hollow, fractured vortex of the looming *Deluge*, dominated by negative hues, is set against the spherical whirl of warmly radiant hope. Yet, as the poetic fragment Turner appended to *Morning After* warns, "Hope's harbinger" — in the guise of the multi-coloured bubble studied by Goethe — is "as ephemeral as the summer fly, Which rises, flits, expands and dies".

For Turner, scientific awe and poetic melancholy coexisted within a single mental spectrum. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Genes for asymmetry easily overruled

The hermit crab *Clibanarius vittatus* inhabits empty gastropod shells and is highly asymmetric. I have found that this asymmetry does not depend on initial possession of a shell, but can be lost within a few moults in the absence of shells, suggesting that the morphology has a genetic component but is environmentally regulated. The loss of dependence on shells has been implicated in the evolution of the nearly symmetrical, fully armoured king crabs and coconut crabs from hermit crab ancestors^{1,2}. The first steps of such a transformation may thus have been much easier than previously thought^{3,4}.

C. vittatus, a common western Atlantic hermit crab, uses shells and has the unarmoured, dextrally coiled abdomen and asymmetrical tail-fan typical of shell-using hermit crabs (Fig. 1a, b). Like most hermit crabs in the family Diogenidae, *C. vittatus* megalopae (the transitional larval stage between pelagic zoeae and benthic adults) are symmetrical but can recognize and use shells⁵.

I reared individuals from hatching to megalopae, giving shells to half of the megalopae⁵, and reared those that metamorphosed into the first postlarval stage for three months with or without shells, dependent on whether they received them as megalopae⁵. Crabs were already asymmetrical at the first juvenile stage, regardless of shell treatment (Fig. 1c, e). Over the next several moults, this initial asymmetry became more exaggerated in shelled juveniles, but weakened in shell-less juveniles (Fig. 1d, f).

In shelled juveniles, the left side of the tail-fan tended to grow in a positively allometric fashion relative to the crab's anterior or carapace, or shield (shield length is the standard metric for hermit crab size⁶), whereas corresponding elements of the right side of the tail-fan showed significant negative allometry (Fig. 1h). In shell-less juveniles, in contrast, left and right elements of the tail-fan showed identical patterns of allometry (Fig. 1g). Left-side elements of shelled and shell-less juveniles grew at the same rate relative to shield length, whereas most right-side elements grew more rapidly in shell-less than in shelled juveniles (Fig. 1g, h).

In several well studied crustaceans that have asymmetrical claws, but are not consistently left- or right-handed, negative feedback between the two sides can be triggered by either internal (genetic or developmental) or external (environmental) forces, whereas control of the direction of the asymmetry is apparently not genetic^{7,8}.

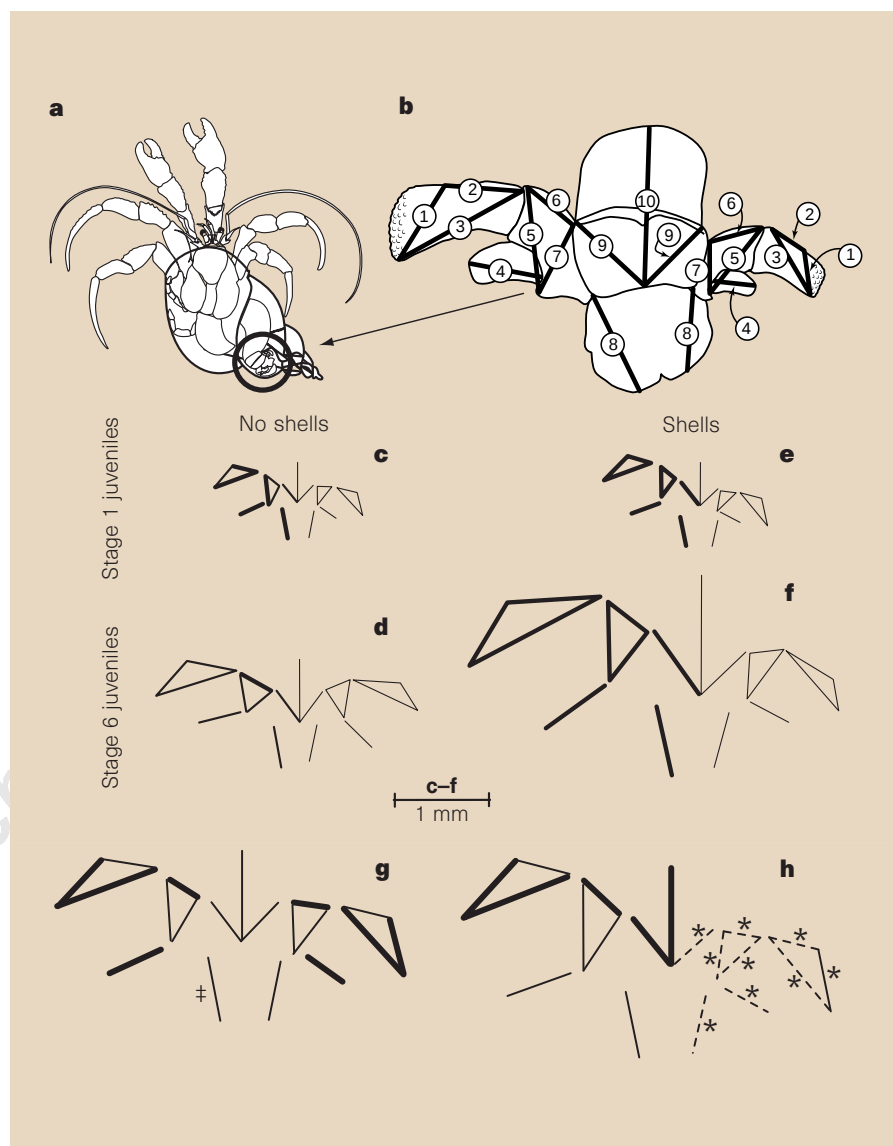


Figure 1 Effects of early shell use on the development of tail-fan asymmetry in *C. vittatus*. **a**, Hermit crab in glass shell; circle and arrow refer to tail-fan *in situ* (after ref. 9). **b**, Tail-fan showing 19 'length elements' (nine left-right pairs of elements plus one unpaired element, the medial sixth abdominal somite). **c-f**, Mean length of elements for juveniles one moult (**c**, **e**) and five moults (**d**, **f**) after metamorphosis, reared without (**c**, **d**) or with (**e**, **f**) shells. Segment thickness of left element indicates whether or not it is significantly longer than the corresponding right element on the same crab (repeated measures ANOVA; thick, $P < 0.005$; medium, $P < 0.05$; thin, $P > 0.05$; $N = 207$). **g**, **h**, Allometric growth in the tail-fan as a function of shell use. **g**, No shells, **h**, shells. Log transformations of each element were regressed against the log-transformed measure of shield length (207 moults). Segment thickness indicates whether the element shows significant ($P < 0.05$) positive allometry (element grew faster than shield; thick solid line), significant negative allometry (element grew more slowly than shield; thin dashed line), or isometry (element and shield grew in size at same rate; thin solid line). Asterisks, elements with significantly lower growth rates; double cross, element with significantly higher growth rate than corresponding elements on both the opposite side of the same treatment and on either side of the other treatment.

In *C. vittatus*, in contrast, it appears that the shell exerts considerable non-genetic control over the degree of asymmetry, whereas the initial right-side reduction in shell-less juveniles suggests that control over the direction of the asymmetry may be genetic. The shell

apparently accentuates the initial tail-fan asymmetry by inhibiting the development of the right side (Fig. 1h). A dextrally coiled shell presumably inhibits right tail-fan development because the crab's abdomen is wrapped around the central column of the shell, with the abdomen's

right side more or less in direct contact with the columella and the left side facing the interior whorl of the shell. Thus, there is room for growth on the left side, but not on the right side⁹.

This mechanism would provide a precise, external means for regulating the degree of asymmetry. The more room available on the left side, the greater the resulting asymmetry, but the initial tailfan asymmetry in shell-less juveniles implies that the direction of asymmetry is determined internally and does not require an external trigger. This raises the possibility of genetic assimilation, an external environmental trigger being replaced by an internal genetic one³.

Not all hermit crab species use dextrally coiled snail shells. Some hermit crabs (for example king crabs and the coconut crab, *Birgus latro*) do not use shelters of any kind as adults¹⁰. The lack of shelter is correlated with increased body symmetry and carcinization (reduced abdomen, with calcified plates, tucked under a broadened carapace^{1,11}). These free-living crabs exhibit some traces of asymmetry, and have been assumed to have evolved from shell-using ancestors^{1,2,11}. However, it has recently been argued that such a transition, requiring both the reversal of a suite of complex morphological characters and the abandonment of a highly adaptive portable shelter, is improbable⁴.

Nonetheless, the fact that many of the distinctive features of the non-shell-using hermit crab genera can be induced in an asymmetrical species within a single ontogenetic cycle, simply by withholding shells, suggests that the developmental constraints against the retention of symmetry in adults are minimal. Intriguingly, abdominal plates, which are almost completely atrophied in adult shell-using hermit crabs, showed signs of redevelopment in shell-less *C. vittatus* juveniles after several moults, implying that even carcinization may be easily triggered.

Alan W. Harvey

Department of Invertebrates,
American Museum of Natural History,
Central Park West at 79th Street,
New York, New York 10024, USA
e-mail: aharvey@amnh.org

1. Cunningham, C.W., Blackstone, N.W. & Buss, L.W. *Nature* 355, 539–542 (1992).
2. Richter, S. & Scholtz, G. *Zool. Anz.* 233, 187–210 (1994).
3. Palmer, A. R. *Proc. Natl Acad. Sci. USA* 93, 14279–14286 (1996).
4. McLaughlin, P. A. & Lemaitre, R. *Contrib. Zool.* (Amsterdam) 67, 79–123 (1997).
5. Harvey, A. W. *Mar. Ecol. Prog. Ser.* 141, 27–36 (1996).
6. McLaughlin, P. A. *Zool. Verh.* 130, 1–397 (1974).
7. Mellon, D. F. & Stephens, P. T. *Nature* 272, 246–248 (1978).
8. Govind, C. K. & Pearce, J. *Science* 233, 354–356 (1986).
9. Brightwell, L. R. *Proc. Zool. Soc. Lond.* 121, 279–283 (1952).
10. Reese, E. S. *Science* 161, 385–386 (1968).
11. Borradaile, M. A. *Brit. Antarctic Terra Nova Exped. 1910 Nat. Hist. Reps. (Zool.)* 3, 111–126 (1916).

A triploblast origin for Myxozoa?

Hox genes, which play key roles in the development of body plans, have been described from a variety of metazoans¹. Here we report the presence of Hox class genes that are typical of triploblasts in Myxozoa, formerly a protozoan taxon. This finding confirms Myxozoa's phylogenetic affinity with the Bilateria and reveals an extreme example of parasitic degeneracy.

The phylum Myxozoa contains more than 1,200 species of endoparasites characterized by a simple cytoplasmic organization, lack of tissue differentiation and multicellular spores containing polar capsules with extrusible polar filaments. Myxozoans have a complex life-cycle, typically alternating between teleost fish and oligochaete worm hosts, in which are formed 'myxosporean' and 'actinosporean' spores.

Both of these spores contain polar capsules, which are strikingly similar in their morphogenesis and mature structure to the nematocysts found in the phylum Cnidaria. The similarity has suggested a close relationship between the myxozoans and cnidarians², and this has been supported by a study of morphological, developmental and 18S ribosomal RNA gene-sequence data³. However, other analyses based on 18S rRNA sequence data have placed the Myxozoa close to the Bilateria⁴.

We attempted to resolve the phylogeny of Myxozoa by characterizing Hox homeodomain sequences of two myxozoans. Degenerate primers that amplify Hox sequences from various Metazoa, including Cnidaria⁵, were used to obtain Hox homeobox sequences for *Tetracapsula bryozoides* and *Myxidium lieberkuehni*, parasitic respectively in the freshwater bryozoan *Cristatella mucedo*⁶ and the pike *Esox lucius*⁴.

Four Hox genes of myxozoan origin were identified from fifteen novel homeobox sequences obtained from spores isolated from host tissue. Eight were discounted as they were also obtained from control host tissue. Some of the remainder may also have been of myxozoan origin but, as they were derived from only one parasite, the evidence was inconclusive.

The four myxozoan Hox inferred amino-acid sequences were compared with two Cnox sequences⁵, using the homeodomain of Antennapedia as a reference sequence (Fig. 1). Sequences Myx-1, Myx-2 and Myx-3 were obtained from both species of myxozoans; they conformed to the consensus sequences of paralogy groups 4, 5, 6 or 7. More specific assignment to these paralogy groups was precluded because informative residues lay outside the ampli-

```

ANTP...HFNRYLTRRRRIEIAHALCLTERQIKI
CNOX1...HF-KKE--T-LSKK-N-S-----
MYX1...-----T--S-----
MYX2...-----R--S-----
MYX3...-----N--G-----
MYX4...L--M--S-E--L--SKSID--D--V--
CNOX2...QN----S-L--Q--AM-D--K-V--

```

Figure 1 Inferred amino-acid sequences of myxozoan Hox homeodomains aligned with Antennapedia (*Drosophila*)¹⁰, and representative cnidarian sequences from *Hydra vulgaris* (Cnox-1)⁵ and *Eleutheria dichotoma* (Cnox-2)⁵. Dashes indicate residues that are identical with those found in the Antennapedia homeodomain.

fied region⁷. Myx-4 was isolated only from *T. bryozoides*. It was not obtained from host material and was confirmed present in parasite genomic DNA by filter hybridization. This sequence belongs to the more-derived Hox-10 paralogy group, hitherto described only from chordates^{7,8}.

None of the myxozoan sequences showed any particular similarity to cnidarian sequences. BLAST searches of the nucleic acid database revealed that the most similar sequences for Myx-1, Myx-2 and Myx-3 are of chordate origin, whereas the most similar sequence to Myx-4 is from *Beroe ovata*, a ctenophore (data not shown). Although formerly classified within the Cnidaria, ctenophores lack true nematocysts. Nielsen interprets their mesoderm formation as indicative of an affinity with the Deuterostomia⁹. If ctenophores indeed have genes of Hox paralogy group 10, this suggests that both groups, formerly identified as closely related to the cnidarians, are triploblastic¹⁰.

This study demonstrates a bilaterian nature for the Myxozoa, but further investigation is required to reveal myxozoan affinities within the triploblasts and explain the extraordinary similarity of myxozoan polar capsules and cnidarian nematocysts. The functional significance of Hox genes in organisms with no distinct anterior–posterior or body axis also remains a mystery.

C. L. Anderson, E. U. Canning

Department of Biology, Imperial College of Science,
Technology and Medicine, London SW7 2AZ, UK
Beth Okamura

Department of Animal and Microbial Sciences,
University of Reading, Whiteknights, PO Box 228,
Reading RG6 2AJ, UK

1. Pendleton, J. W., Nagai, B. K., Murtha, M. T. & Ruddle, F. H. *Proc. Natl Acad. Sci. USA* 90, 6300–6304 (1993).
2. Lom, J. & de Puytorac, P. *Protistologica* 1, 53–65 (1965).
3. Siddall, M., Martin, D. S., Bridge, D., Desser, S. S. & Cone, D. J. *Parasitol.* 81, 961–967 (1995).
4. Schlegel, M. et al. *Archiv. Protistkd.* 147, 1–9 (1996).
5. Kuhn, K., Streit, B. & Schierwater, B. *Mol. Phylogenet. Evol.* 6, 30–38 (1996).
6. Canning, E. U., Okamura, B. & Curry, A. *Fol. Parasitol.* 43, 249–261 (1996).
7. Sharkey, M., Graba, Y. & Scott, M. P. *Trends Genet.* 13, 145–151 (1997).
8. García-Fernández, J. & Holland, P. W. H. *Nature* 370, 563–566 (1994).

9. Nielsen, C. *Animal Evolution. Interrelationships of the Living Phyla* 10–317 (Oxford Univ. Press, Oxford, 1995).
10. McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. *Nature* **308**, 428–433 (1984).

Coral provides way to age deep water

We propose a new method for computing the ratio of the isotopes carbon-14 to carbon-12 in deep water from the past, and for testing the results derived from the normal method of age difference between benthic and planktic foraminifera in deep-sea sediments¹.

Our method involves measuring the ages of solitary corals that lived in the deep sea. The thorium/uranium ratio in the coral yields its true age, whereas the carbon-14 signature represents the age of the coral plus the age of deep-water dissolved inorganic carbon, from which it is formed. Therefore, subtracting the thorium/uranium age from that of carbon-14 yields the age of the deep water at the time the coral was formed. Application of this method to a deep-sea coral from a water depth of 2,306 metres in the Atlantic Ocean yielded a deep-water ventilation age (the time since it was last 'ventilated' to the air at the ocean surface) of about 1,200 years during the meltwater pulse-1A (ref. 2), a rapid phase of ice-melting during the last deglaciation.

Carbon-14 (¹⁴C) is useful as a tracer because, unlike other nutrients, its record in biological carbonates is unaffected. Other tracers such as cadmium/calcium or $\delta^{13}\text{C}$ may be affected by pH, the geochemistry or the environment where the forams lived. Deep solitary corals are superior to foraminifera for reconstructing the circula-

tion of the deep sea, because these corals are unaffected by bioturbation^{3,4}. Moreover, protactinium-231/uranium disequilibrium dating by thermal ionization mass spectrometry can be used to corroborate the coral's 'true age' derived from its thorium/uranium ratio (Th/U).

We analysed a solitary ahermatypic coral sample that was collected from gravity box core GeoB 1503-1 (02° 18.7' N, 30° 38.8' W) taken at 2,306 metres depth during the cruise of the research vessel *Meteor* 16/2 (ref. 5). The coral lay at a depth of 17 cm in the core, in a layer rich in pteropods. This layer probably corresponds to the deep-sea carbonate-preservation spike that has been observed in the Atlantic Ocean, centred at an age of 13,500 years (¹⁴C) and with a maximum at 14,000 years (ref. 6). The coral species was classified as *Caryophyllia ambrosia* by H. Zibrowius (Marseilles).

The initial, age-corrected activity ratio for uranium-234/uranium-238 (Table 1) was found to be 151.9 ± 3 (delta notation). This comes very close to the ratio of present-day sea water, 144 ± 4 (ref. 7), suggesting that diagenetic alteration of the sample has been negligible. The 'bottom' sample taken from the lower section had a very low ²³²Th content of 0.7 nanograms per gram. This, together with the very good agreement of the thorium and protactinium radioactive clocks, suggest that the cleaning procedure was successful and that the age derived from the Th/U-method is reliable.

The difference of ages between the 'bottom' and 'top' samples suggests that the coral lived less than 200 years. The 'true age' of the solitary coral — 14,100 years — can be converted into a terrestrial ¹⁴C age of 12,047 years by applying the ¹⁴C-calibration curve⁸. The age difference between the ¹⁴C age and 'true age' of the 'bottom' section then amounts to $1,083 \pm 125$ years ($\pm 1\sigma$),

determined as $13,530 \pm 75$ minus $12,447 \pm 100$ years, assuming a reservoir effect of 400 years. This method is similar to the one applied to forams.

However, 14,100 years ago the atmospheric $\Delta^{14}\text{C}$ (¹⁴C/¹²C relative to an oxalic standard) amounted to 221.1 ± 14 ‰ and was decreasing at a rate of approximately 30‰ per 1,000 years⁸. Thus the deep water bathing the deep-sea coral equilibrated with an atmosphere that had a ¹⁴C/¹²C ratio, C^{past} , enriched in ¹⁴C compared with the atmosphere synchronous with the coral growth.

The relationship between a non-reservoir-corrected ¹⁴C age and the measured ¹⁴C/¹²C ratio, C^{meas} , applying the conventional mean life of ¹⁴C, is

$$C^{\text{meas}} = C^{\text{stand}} \cdot e^{-(C^{\text{meas}}/8033)} \quad (1)$$

where C^{stand} is the ¹⁴C/¹²C ratio of modern (pre-1950) carbon. The present-day ¹⁴C/¹²C ratio in the coral can then be approximated⁹, assuming 'closed system' radioactive decay during the trajectory of the water mass from when it was last 'ventilated' to the atmosphere at the ocean surface, to the point of the coral sampling (ref. 7), as

$$C^{\text{meas}} = C^{\text{past}} F^{\text{ML,past}} e^{-[(\text{Th/U-age}+T)/8266]} \quad (2)$$

where $F^{\text{ML,past}}$ takes into account the reservoir effect in the mixed layer (at present about 0.95) and T is the ventilation age. The ventilation age can be computed by applying equations (1) and (2) as:

$$T = -8266 \ln(C^{\text{stand}}/C^{\text{past}} 1/F^{\text{ML,past}} \times e^{-(C^{\text{meas}}/8033)}) - \text{Th/U-age [years]} \quad (3)$$

With values of F^{ML} of 0.95 and of C^{past} between 1.23 and 1.27, the present-day value yields ventilation ages between 1,077 and 1,342 years. This long ventilation time of deep water 12,000 ¹⁴C-years ago shows that the null hypothesis, that water-mass mixing ratios did not change and older surface water was supplied in similar proportions to today, is unlikely. For example, if an older reservoir age of surface water in the past is assumed, such as an F^{ML} of about 0.91 (ref. 10), a ventilation age of 855 years is obtained. A ventilation age of 350 years would require a F^{ML} as low as 0.87.

The long ventilation time was therefore caused either by sudden homogenization of a stratified water column^{11,12} when older deep water with properties similar to those of Southern Component Water (a mixture of Weddell Sea bottom water and ambient circumpolar water) was mixed vertically, or was due to a larger contribution by Southern Component Water coincident with the meltwater pulse-1A (ref. 13).

The latter explanation agrees with predictions made by numerical models for periods of lower salinity in areas of deep-water formation in the North Atlantic¹⁴.

A. Mangini, M. Lomitschka, R. Eichstädter, N. Frank, S. Vogler

Heidelberger Akademie der Wissenschaften,
INF 366, 69120 Heidelberg, Germany
e-mail: mg@uphys1.uphys.uni-heidelberg.de

Table 1 Ageing of coral

		'Bottom'	'Top'
²³⁸ U	[μg g ⁻¹]	4.186±0.016	3.847±0.008
²³⁴ U/ ²³⁸ U	[initial]	1.152±0.003	1.154±0.004
²³² Th	[ng g ⁻¹]	0.677±0.004	3.590±0.024
²³⁰ Th	[pg g ⁻¹]	9.534±0.066	8.866±0.220
²³¹ Pa	[pg g ⁻¹]	0.352±0.01	Not measured
Th/U-age	[yr BP]	14,100±180	14,230±420
Pa/U-age	[yr BP]	14,050±500	Not measured
¹⁴ C-age	[¹⁴ C yr BP±1σ]	13,530±110	1st fraction 12,870±110 2nd fraction 13,000±100 3rd fraction 13,210±110 Remainder 13,420±110

Coral was cut into two sections perpendicular to the growth pattern ~2 cm from the bottom. The lower section (~4 g, 'bottom') was homogenized, and pretreated Mn-coatings¹⁵ and samples were taken. Th-, Pa- and U-isotopes were measured by TIMS¹⁶. All values $\pm 2\sigma$. ¹⁴C dating was performed on the top section (first fraction through to 'remainder') which had not been pre-cleaned. The increasing ages in the different fractions represent a cleaning from embedding sediment rather than an alteration of the coral surface. We also performed a further Th/U dating on the 'top' part. Pa was measured with a two-filament technique at an ionization temperature of more than 1,900 °C. The amount of added ²³³Pa spike was about 95 fg, yielding count-rates of about 10 s⁻¹ for ²³³Pa and 100 s⁻¹ for ²³¹Pa. The blank contribution of ²³¹Pa was less than 0.5% of the sample. Calibration of the ²³³Pa spike was performed on a low-level HPGe well-detector using the photopeak at 312 keV with a branching ratio of 0.386. The efficiency of the γ-detector at this energy was $6.41 \pm 0.13\%$.

G. Bonani, I. Hajdas

Institut für Mittelenergie-Physik,
ETH-Hönggerberg, CH-8093 Zürich, Switzerland

J. Pätzold

Fachbereich Geowissenschaften,
Universität Bremen, 28359 Bremen, Germany

1. Broecker, W. S., Peng, T. H., Trumbore S., Bonani, G. & Wölfl, W. *Global Geochem. Cycles* 4, 103–107 (1990).
2. Bard, E., Hamelin, B. & Fairbanks, R. G. *Nature* 346, 456–458 (1990).
3. Broecker, W. S., Mix, A., Andree, M. & Oeschger, H. *Nucl. Instr. and Meth. B5*, 331–339 (1984).
4. Peng, T. H. & Broecker, W. S. *Nucl. Instr. Meth. B5*, 346–352 (1984).
5. Schulz, H. D. *et al. Ber. Fachber. Geowissenschaften, Univ. Bremen* 19, (1991).
6. Berger, W. H. *Nature* 269, 301–304 (1977).
7. Chen, J. H., Edwards, R. L. & Wasserburg, G. J. *Earth Planet. Sci. Lett.* 80, 241–251 (1986).
8. Stuiver, M. & Reimer, P. J. *Radiocarbon* 35, 215–230 (1993).
9. Adkins, J. F. & Boyle, E. A. *Paleoceanography* 12, 337–344 (1997).
10. Bard, E. *et al. Earth Planet. Sci. Lett.* 126, 275–287 (1994).
11. Sarnthein, M. *et al. Paleoceanography* 10, 1063–1094 (1995).
12. Oppo, D.W. & Fairbanks, R. G. *Earth Planet. Sci. Lett.* 86, 1–15 (1987).
13. Bard, E. *et al. Nature* 382, 241–244 (1996).
14. Fichet, T., Hovine, S. & Duplessy, J. C. *Nature* 372, 252–255 (1994).
15. Lomitschka, M. Diplomarbeit Inst. für Umweltphysik, Univ. Heidelberg (1996).
16. Bollhöfer, A., Eisenhauer, A., Frank, N., Pech, D. & Mangini, A. *Geol. Rundsch.* 85, 577–585 (1996).

Contradictions of slate formation resolved?

Slate is formed from clay-rich mud in response to tectonic stress. It has been studied for more than 150 years and was among the first geological features to be analysed on the microscopic scale¹. Early observations² recognized the importance of mica-ceous minerals to the splitting of slate into thin sheets, whereas current hypotheses for slate formation emphasize either mechanical processes (grain rotation and grain kinking) or chemical processes (grain dissolution and new growth). Despite a vast body of work, no single scenario incorporates these seemingly contradictory mechanisms^{3,4}. Here we offer a unifying model

that views the mechanisms as a function of the combined thermal and strain energy of the system.

This model for slate cleavage has arisen from recent electron microscopy and X-ray work on slates from different, well constrained environments. The model focuses on the respective roles of phyllosilicates of depositional origin (detrital grains) and those formed during diagenesis and metamorphism (authigenic grains). Optical and scanning electron microscopy can be used to resolve detrital grains, which are about 10 micrometres in size, whereas transmission electron microscopy is needed for authigenic grains, which are about 10 nanometres. To measure the crystallographic orientation of all phyllosilicates in proportion to the volume of a particular population in a specific orientation, X-ray texture goniometry is used.

The Rhayader district of the Welsh basin in the United Kingdom preserves a sequence of progressively cleaved mudstones that range in metamorphic grade from diagenetic zone (zeolite facies) to epizone (lower greenschist facies) over a distance of about 10 kilometres⁵. In the lower-grade portion of the area, the detrital mica and chlorite progress from bedding-parallel to cleavage-parallel orientations⁶. Electron microscopy shows variably oriented authigenic mica (Fig. 1a). In higher-grade samples, detrital phyllosilicates remain parallel to bedding, whereas fine-grained, relatively strain-free, metamorphic phyllosilicates are parallel to cleavage (Fig. 1b). Transitional bedding-cleavage orientations are rarely observed.

In contrast, mudstones at Lehigh Gap, Pennsylvania, in the United States, show gradual cleavage development over a distance of about 100 metres in response to a local strain gradient under constant (anchizone) metamorphic grade⁷. X-ray texture goniometry and electron microscopy show that mechanical rotation of detrital grains dominates cleavage formation in the low-strain region, whereas dissolution of primarily authigenic phyllo-

silicates and neocrystallization produced cleavage in the high-strain region^{8,9}.

Grain deformation and rotation of detrital phyllosilicates dominate cleavage formation in the lowest-strain and lowest-grade rocks, from Lehigh Gap and Rhayader, respectively. But the higher-strain and higher-grade rocks of these areas show that dissolution and neocrystallization of fine-grained, authigenic phyllosilicates are primarily associated with cleavage formation. Evidence for mechanical deformation is either obliterated by subsequent neocrystallization in the high-strain samples of Lehigh Gap, or was developed only sporadically in the high-grade samples from Rhayader.

Despite distinctly contrasting environmental conditions, the two rock sequences show the same progression of deformation processes. In both the Rhayader samples that developed under a metamorphic/thermal gradient and the Lehigh Gap samples that developed under a strain gradient, we find a progressive change from dominantly mechanical processes in detrital grains to chemical processes in authigenic grains.

This indicates that the roles of thermal and strain energy are indistinguishable during cleavage formation. Also, assuming that the total energy defines the dominant deformation mechanism, the preference of mechanical processes in 'low-grade' and 'low-strain' regions indicates that mechanical work requires less energy than chemical work.

The mechanism that predominates in slate cleavage-forming is thus a function of the combined effects of thermal and strain energy in rocks. The contribution of either source to the total energy of the system is complementary and interchangeable. Mechanical processes (grain kinking and rotation) are favoured in relatively low-energy environments, whereas chemical processes (grain dissolution and neocrystallization) are favoured in relatively high-energy environments.

Ben A. van der Pluijm, Nei-Che Ho,

Donald R. Peacor

University of Michigan, Ann Arbor, Michigan, USA

Richard J. Merriman

British Geological Survey, Keyworth,
Nottingham, UK

e-mail: vdpluijm@umich.edu

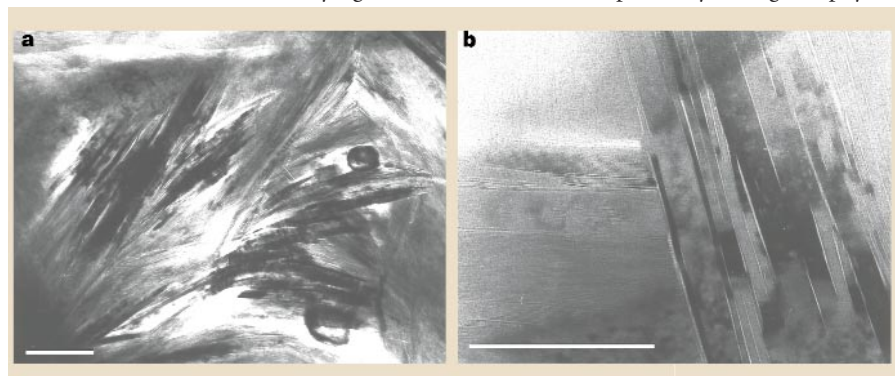


Figure 1 Transmission electron micrographs from Welsh basin rocks. **a**, Irregularly oriented and bent authigenic mica from low-anchizone slates. **b**, Recrystallized, cleavage-parallel metamorphic mica from epizone slates. Bedding in both images is horizontal, whereas cleavage is nearly vertical. Scale bar, 100 nm.

1. Sorby, H. C. *Edinb. New Phil. J.* 55, 137–148 (1853).
2. Darwin, C. *Geological Observations on South America* (1846).
3. Knipe, R. J. *Tectonophysics* 78, 249–272 (1981).
4. Engelder, T. & Marshak, S. J. *Struct. Geol.* 7, 327–343 (1985).
5. Merriman, R. J., Roberts, B. & Hiron, S. R. *BGS Tech. Report WG/91/16* (1992).
6. Ho, N.-C., Peacor, D. R. & van der Pluijm, B. A. *J. Struct. Geol.* 18, 615–623 (1996).
7. Epstein, J. B. & Epstein, A. G. *Fieldguide* 132–205 (Rutgers Univ. Press, 1969).
8. Lee, J. H., Peacor, D. R., Lewis, D. D. & Wintsch, R. P. *J. Struct. Geol.* 8, 767–780 (1986).
9. Ho, N.-C., Peacor, D. R., & van der Pluijm, B. A. *J. Struct. Geol.* 17, 345–356 (1995).

American voyeurs in Paris

Against the Spirit of System: The French Impulse in Nineteenth-Century America

by John Harley Warner

Princeton University Press: 1998. Pp. 510.
\$49.50, £37.50

To See with a Better Eye: A Life of R. T. H. Laennec

by Jacalyn Duffin

Princeton University Press: 1998. Pp. 456.
\$49.50, £35

W. F. Bynum

For much of the twentieth century, doctors from all over the world have appreciated the value of the qualification familiarly dubbed the BTA: been to America. But in the first half of the nineteenth century, Paris was the Mecca of medicine, and thousands made the pilgrimage. They came from all over Europe, but no country was better represented than the United States. Some were simply medical tourists, but many more stayed in Paris for months or years, immersing themselves in the sights, sounds and smells of the clinics, morgues, lecture halls and dissecting rooms.

The phenomenon itself is well known: an episode in the stately movement of the medical avant-garde from Padua to Leiden to Edinburgh before Paris, and then on to Vienna and Berlin, Boston, New York and Baltimore. John Harley Warner's achievement is to have taken a historical commonplace and shown how richly complex it actually was.

His evidential base is staggering: manuscript letters, diaries and reminiscences from more than 70 US repositories, and a couple of dozen from Britain, Canada and Paris. Some material has been used before, of course.



PIERRE CHARLES ALEXANDRE LOUIS

Clinical paragon: Pierre Louis.

James Jackson Sr published many of his son's letters home, as a poignant tribute to his talented offspring who died shortly after returning to Boston, aged 24; and Sir William Osler immortalized John Young Bassett, the 'Alabama student' of his famous essay of that name. But much of Warner's archival digging is new and all of it is newly interpreted.

René Laennec, Pierre Louis, Gabriel Andral, Guillaume Dupuytren and the other leading lights of the Parisian medical scene had international reputations. ("At Paris I am to study under the influence of men whose aim is science — at London, gain" wrote James Jackson Jr.) But, as Warner shows clearly, it was the patients, living and dead, as much as the teachers that made students choose Paris over London or some other urban centre.

They went to touch the living and dissect the dead: nowhere were there so many tractable bodies on offer. For two francs, pregnant women were prepared to be examined by a dozen students, or deliver their babies in full view of 50. For 30 francs, one could dissect, along with 200 colleagues, for six months at the amphitheatre at Clamart, whence unclaimed bodies from all the hospitals were taken each morning. For a small fee, an *interne* could be hired for private tuition who could unlock the wards whenever he wished. And lectures and large-group hospital rounds were free to foreigners.

There was a down side, of course: time, expense, loneliness and danger, sometimes mortal, from fevers in the clinics or dissection wounds in the morgues. And most students came with little French, so language tutorials were a first priority. Some began eventually to write their diaries in schoolboy French and a few even used it in their letters back home. Because the letters and diaries were produced for family consumption, they contain little about Paris low-life, carousing or brothels. But there was pretty unanimous agreement that the French were often morally loose and that French doctors were on the whole brilliant diagnosticians and inept healers.

As a consequence, Americans selected what they thought best from the French: a clinical science based on the bedrock of observation and fact, and purged of speculation and 'system'. Warner takes his title — *Against the Spirit of the System* — from the campaign waged in mid-century America by the cohort of doctors who had experience of French medicine or were inspired by its precepts. "And by science in medicine, we mean large experience, not theory", wrote one doctor. For most, the man who perfectly embodied this kind of science was Pierre Louis (1787–1872), with his *méthode numérique*



Medical symbol: René Laennec.

and commitment to long years of careful bedside observation. His band of devoted US disciples needed to be doubly devoted, as Louis never bothered to learn English.

This interplay of French and US themes makes Warner's book exceptionally rich, for Americans in Paris noticed things the natives took for granted. At the same time, what Warner calls the French 'impulse' became a major force in American medical culture, and the closing chapters analyse the transplantation and adaptation of French ways of doing things within the very different environment of the New World. Only a few Americans dared to embrace totally the 'expectant' therapeutics of the French, as this played into the hands of critics who argued that to stand by and watch a disease take its course was callous; it was in any case unlikely to be successful within the US medical marketplace. Nevertheless, Parisian ideals helped to shape mid-century medical science in the United States and provided a platform for professional reform.

Although Louis was the clinical paragon for the Americans in Paris, the man who more than any other came to symbolize this medical epoch was René Laennec (1781–1826), inventor of the stethoscope. Foreigners were beginning to frequent his lectures in the 1820s, but he died before the larger student influx of the following decades, and his influence was mostly indirect. Nevertheless, auscultation was widely held to have transformed diagnosis from an art into a science, albeit one that still relied on the confirming evidence of the mortuary.

Despite Laennec's iconic status in the medical pantheon, he has been badly served

biographically, even by the French. A two-volume study by Alfred Rouxeau early in this century had the virtue of reproducing a good deal of the Laennec family correspondence, but Rouxeau sanitized Laennec's sometimes difficult personality (and awkward family) and seemed vaguely embarrassed by his final medical statement, the unpublished lectures he delivered at the Collège de France from the early 1820s. These show Laennec the polemicist at his starkest; they also contain his final thoughts on vitalism, ideas which were counter to the dominant organicism of his contemporaries. Jacalyn Duffin speculates in *To See with a Better Eye* that Laennec's nephew and executor did not edit the lectures for publication for fear that they might be seen as unworthy of the great clinician and pathologist, an aberration consequent on his Roman Catholicism.

To her great credit, Duffin takes on the whole Laennec. Her elegant biography wonderfully exploits the surviving Laennec archives and offers a rounded portrait of this complex and highly talented doctor. That he was socially, politically and religiously conservative does not make his contribution to medicine any the less revolutionary. Duffin guides us expertly through the 1819 and 1826 editions of Laennec's *De l'auscultation médiate*, historically dissecting many of the cases on which Laennec based his new technique of stethoscopy. More generally, her analysis of the vibrant medical scene in Paris provides ample evidence of why students flocked there.

"I live among the dead and the dying," Laennec wrote to his cousin in 1810. On the bodies of those dead and dying were laid the foundations of modern clinical medicine. These two fine monographs complement each other, and allow us to appreciate exactly why the *clinique* left such a powerful imprint on the nineteenth-century medical mind. □

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

Attention!

A paperback on attention has just been published that unfortunately arrived too late for inclusion in Stuart Sutherland's review on this page.

In *Attention*, Harold Pashler, the author of one of the books assessed by Sutherland, gathers together essays by 12 leading international researchers examining different facets of contemporary research in the subject. The volume is divided into two sections, one dealing with psychological research such as visual search, dual-task interference and attentional bottlenecks, and the other with modern approaches to neural-network modelling and the effects of brain damage on attention. The tutorial-style chapters make the book ideal for students. Psychology Press, £14.95.

Feature selection

The Psychology of Attention

by Harold E. Pashler

MIT Press: 1997. Pp. 471. \$45, £38.50

The Psychology of Attention

by Elizabeth A. Styles

Psychology Press: 1997. Pp. 259. \$64.95,

£44.95 (hbk); \$37.95, 14.95 (pbk)

Stuart Sutherland

Over the past 50 years, the sheer ingenuity displayed by psychologists working on attention rivals if it does not exceed that of cosmologists studying black holes. Indeed, there is a similarity in their results — after many thousands of experiments, we know only marginally more about attention than about the interior of a black hole. As Harold Pashler ruefully remarks of one set of experiments, "these results create a rather unappetising stalemate".

As in a Punch and Judy show, as soon as one investigator conclusively proves a thesis, another pops up and knocks him down. Donald Broadbent found that if a subject repeated a string of words presented to one ear, he heard none of a string simultaneously presented to the other ear. He concluded that there are peripheral filters that prevent an unattended message being centrally processed, whereupon Judy, in the shape of Neville Moray, promptly clobbered him by showing that if the subject's own name was presented on the unattended ear, it was heard on about a third of occasions.

There is much further evidence that messages can be unconsciously processed centrally before being rejected. A digit is more readily found in an array of letters than in an array of digits. In a clever experiment, it was shown that this was not because digits differ systematically in shape from letters. If subjects are asked to find the letter 'O' among distracting digits, they are faster than if asked to find the number zero in the same array of digits. Attentional selection of one of two messages is occurring at a central semantic level.

Much of the work on attention attempts to discover which processes go on in parallel and which serially. Common sense will probably tell you that at least some processing must be sequential since you cannot for the most part do two things at once. But you can. Subjects can be taught to read while writing to dictation. After much practice they do both tasks together as efficiently as either on its own.

Pashler claims there is a bottleneck through which the processing of both tasks must pass, but that with practice it is possible to program in advance a string of responses (although he does not make the comparison, this corresponds to 'chunking' in sensory domains). Hence, by devoting less processing time to selecting each response it is possible to interweave the processor between the two tasks. The automation of a task is a fascinating topic.

There would certainly appear to be a bottleneck in consciousness: practised motorists may drive for miles while chatting animatedly without any awareness of their driving.

Anne Treisman discovered recently that when a simple form (for example, a bar) differs from an array of distracting forms in only one elementary feature (such as orientation) it 'pops out' at the observer. It is located immediately, while in other circumstances all the shapes have to be searched serially to find a specified target. This result excited considerable interest among neurophysiologists, as the features for which the effect held seemed to correspond to those firing 'feature detectors' in the early stages of visual processing.

But this conclusion was quickly refuted by another cunning experiment in which it was found that a systematic difference in perceived size could cause the pop-out effect even when the retinal size of the distractors was varied because they were at different distances.

Both Pashler and Elizabeth Styles make valiant and on the whole successful attempts to clarify the literature on attention, as is attested by the fact that material found in one quite frequently does not appear in the other. Pashler's is slightly the better book. It is broader, covering central mechanisms of attention more thoroughly; sometimes, though too seldom, it asks why a given mechanism should exist; and it is more thoughtful, making some attempt to connect scientific findings with the demands of everyday life.

Both authors have a staggering mastery of the field, but neither book conveys all the important information on attention. The complexity of the theories and experiments and of the relation between the two is such that the books are not easy. Do not attempt to read them while taking dictation — they require your undivided attention. □

Stuart Sutherland is at the Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

After Simpson

Classification of Mammals: Above the Species Level

by Malcolm C. McKenna and Susan K. Bell

Columbia University Press: 1997. Pp. 631.

\$175, £140

Jean-Louis Hartenberger

In 1662 and 1663, John Ray, the founder of natural history in England, made a journey to the main European universities. I like to imagine that the stop he made in my city, Montpellier — where he met Niels Stensen (also known as Nicolaus Steno) and Martin Lister, with whom he dissected various animals — had some influence on his subsequent writing of *Synopsis methodica animalium quadrupedum et serpentini generis* (1693), one of the



Chilling out: polar bears relaxing, from *Polar Dance: Born of the North Wind* (Images of Nature, \$65) with photography by Thomas D. Mangelsen and text by Fred Bruemmer.

great landmarks of vertebrate zoology.

After the elaboration of a hierarchical method of biological classification, and mainly on the basis of Ray's work, Carl Linnaeus gave birth in 1758 to the concept of Mammalia in which *Homo* is considered to be a member of Primates. From that time we have been more concerned with classifying and systematizing mammals than we have other living organisms.

Following Georges Cuvier, numerous tentative schemes for classifying mammals have been made in which fossils ("les espèces perdues") were considered along with extant species. Of all the pre-Darwinian classifications, my preference goes to that of Ducrotay De Blainville (1834), for two main reasons: he gave less weight to adaptive characters for defining higher taxonomic levels, emphasizing instead 'deep' or 'fundamental' ones; and he arranged some fossil genera in linear series to form some kind of 'natural transition' for filling the gaps among living animals. In addition, the rodents, in which I have a particular interest, are very well grouped by de Blainville.

After Darwin, palaeomammalogists generally recognize that the most comprehensive and documented proposals in the field of mammalian systematics came from the American Museum of Natural History in New York. At the beginning of the century, William King Gregory (1908) offered a brilliant and still valuable synthesis of mammalian relationships. In the 1940s, George Gaylord Simpson worked on the task for

many years, before and after the publication in 1945 of his classification of mammals. In many ways, more recent authors consider that their own systematic work expands on that of Simpson. We can therefore speculate that the new classification of Malcolm C. McKenna and Susan K. Bell, an extensive revision of Simpson's classic work, will be in use for the next 50 years.

This new work embraces no less than 5,000 genera (of which 79% are extinct), distributed in 425 families (70% are extinct) and 46 orders (22 are extinct). In Simpson's classification, 'only' 2,864 genera (932 were living) were distributed in 257 families, of which 'only' 54% were extinct. So in the past 50 years there have been as many new fossil discoveries as there were in the time from Cuvier to Simpson. If we consider Primates, favourites of (too) many colleagues, the point is even more striking: to the 82 genera of 1945, 135 more have been added.

So when faced with these numbers and keeping in mind how 'bushy' is the plethora of taxa and nomenclature proposed in thousands of papers, one can understand the attitude of many investigators who were asked to give advice about the aims of our colleagues at the American Museum of Natural History: most of us thought this venture was some sort of Sisyphean task. But computers have helped to handle this enormous database, and McKenna, Bell and their colleagues were convinced of the need for a new classification. They make a great attempt to communicate in words their

depiction of a phylogenetic branching and descent with modification of the class Mammalia. To accomplish this, they organize the classification using 25 taxonomic hierarchical levels, ten more than Simpson used. For this, they consider that phylogenetic analyses (cladistics) are more of a tool than a subject of study.

The results in many cases are provocative, and no doubt many modifications will be suggested in the future. In my opinion, the book delivers two messages that are invitations for future research. The first (reading between the lines) is that characters are not black and white, and we need to do more than looking at previous papers and compiling lists of characters for cladistic analyses. What we need is an understanding of characters and their biological significance; undoubtedly we need new neontological studies, particularly in the developmental field, at all available levels. The second message addresses young palaeontologists: if they want to shake the tree of mammals, as depicted in the McKenna and Bell classification, the best way to do this is to search in the field for new species and genera. In the next 50 years they have to find at least 2,000 more genera, such that about 90% of known families will be extinct, for constituting the skeletal framework of future mammalian classifications.

In the meantime, the classification of extant and fossil mammals by McKenna and Bell will be the classic reference work for all investigators interested in mammalian evolutionary biology. □

Jean-Louis Hartenberger is in the Laboratoire de Paléontologie, Université Montpellier II, F-34095 Montpellier cedex 5, France.
e-mail: hartenjl@isem.univ-montp2.fr

Jargon and icons

Oxford Dictionary of Biochemistry and Molecular Biology

edited by A. D. Smith, S. P. Datta, G. Howard Smith, P. N. Campbell, H. A. McKenzie and R. Bentley
Oxford University Press: 1997. Pp. 740.
£34.95, \$60

Christopher Surridge

One of the most daunting aspects of modern biology is the vast amount of jargon shrouding it. There is an ever-expanding vocabulary, and hardly a week passes without the coining of a name or acronym in *Nature* alone. The *Oxford Dictionary of Biochemistry and Molecular Biology* attempts to provide a guide in this linguistic jungle for novice and experienced traveller, indeed for anyone who wants to know the difference between IGF, IgG and IGT (insulin-like growth factor, immunoglobulin G and impaired glucose tolerance).

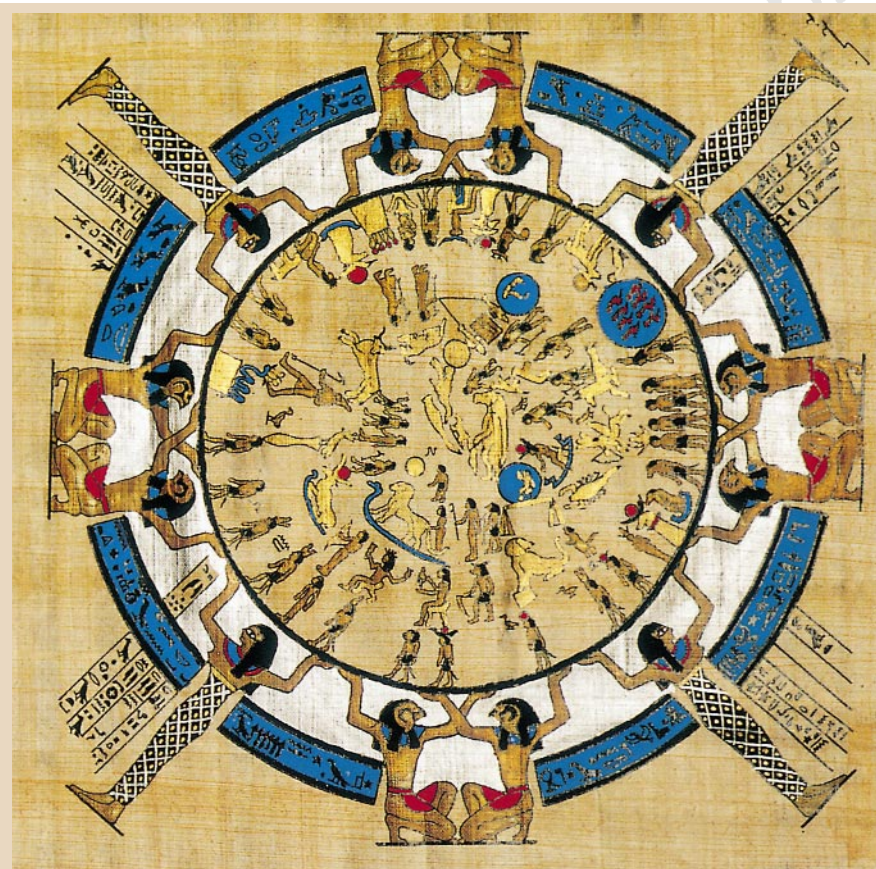
Luckily the editors — or lexicographers as they prefer to be called, allowing for inclusion of Dr Johnson's definition of a lexicographer as "a harmless drudge" — have had the good sense to realize that a mere list of definitions, although useful, would fail to acknowledge the countless scientists who have formed the field, many leaving their names behind in the process. So they tell us that, whereas the Lewis acids were named after the US chemist G. N. Lewis (1875–1946), Lewis antigens derive quite literally from Mrs H. D. G. Lewis in whom they were first identified. A large number of short biographical entries are also included for the more important personalities in the field; well, Nobel laureates anyway.

Another aspect that raises the dictionary above a simple list of terms and meanings comes from the inclusion of structure diagrams. These have long been vital pieces of information for the modern biologist, for whom form and function are intimately

linked. Some pages may look like a cross between the Aldrich catalogue and a seven-year-old's *My First Picture Dictionary*, but then few of us could guess the relative molecular mass of the reserpine in the fridge, even if we knew it to be (3b,16b,17a,18b,20a)-11,17-dimethoxy-18[(3,4,5-trimethoxybenzoyl)oxy]yohimban-16-carboxylic acid methyl ester.

Along with the main text, the appendices provide a good grounding in the mysteries of bioinformatics, nomenclature conventions and so on, making this a useful first point of reference. Of course, in such a fast-growing field, omissions are unavoidable — anyone wishing to know the difference between MAD (multiwavelength anomalous diffraction) and Mad (Mothers against decapentaplegia) must search elsewhere — but any copy of this book that finds itself on a research laboratory's bookshelf will be dog-eared and coffee-stained in a matter of months.

Christopher Surridge is an assistant editor of *Nature*.



Making sense of the stars

The ancient Egyptians combined star-gazing with mythology, as this painting on papyrus from a Ptolemaic temple demonstrates. As well as showing the signs of the zodiac of Dendara, it includes *sa'mut* (the hippopotamus with a crocodile on her back), *Mesketiu* (the bull) and *Selket* (the scorpion goddess). But it was the Greeks who laid the foundations of astronomy as we know it. The invention of the telescope

provided the next great leap forward, and advances in the past few years have given us a much more detailed view of the heavens. The development of astronomy from the early astrology of the Babylonians in 5000 BC to Hubble and beyond are chronicled by Robert Wilson in *Astronomy Through the Ages* (Taylor and Francis/ Princeton University Press, £19.95, \$29.95).

New editions

Ozone Diplomacy: New Directions in Safeguarding the Planet

by Richard Elliot Benedick
Harvard University Press, \$42.95, £26.50 (hbk),
\$18.95, £12.50 (pbk)

Modern Mycology, 3rd edn

by J. W. Deacon
Blackwell Science, £21.50 (pbk)

Principles of Population Genetics, 3rd edn

by Daniel L. Hartl and Andrew G. Clark
Sinauer, £39.95 (hbk)

Elements of Petroleum Geology, 2nd edn

by Richard C. Selley
Academic, \$74.95 (hbk)

"An excellent, very readable text for final-year undergraduates", wrote *Nature's* reviewer of the first edition of 1985.

Epidemiology for the Uninitiated, 4th edn

by D. Coggon, G. Rose and D. J. P. Barker
BMJ Books, £9.95 (pbk)

Agronomy of Grassland Systems, 2nd edn

by C. J. Pearson and R. L. Ison
Cambridge University Press, £55, \$80 (hbk),
£19.95, \$32.95 (pbk)

Plant Variation and Evolution, 3rd edn

by D. Briggs and S. M. Walters
Cambridge University Press, £65, \$80 (hbk),
£22.95, \$34.95 (pbk)

A Primer of Drug Action: A Concise, Nontechnical Guide to the Actions, Uses, and Side Effects of Psychoactive Drugs, 8th edn

by Robert M. Julien
W. H. Freeman, \$21.95, £16.95 (pbk)

Principles and Practice of Soil Science: The Soil as a Natural Resource, 3rd edn

by R. E. White
Blackwell Science, £21.50, \$42.95 (pbk)

Introduction to Physical Chemistry, 3rd edn

by Mark Ladd
Cambridge University Press, £65, \$100 (hbk),
£22.95, \$44.95 (pbk)

Images of the Earth: Essays in the History of the Environmental Sciences, 2nd edn

edited by L. J. Jordanova and Roy Porter
British Society for the History of Science, £14.95,
\$27 (pbk)

The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*

Gerard Deckert^{*†}, Patrick V. Warren^{*†}, Terry Gaasterland[‡], William G. Young^{*}, Anna L. Lenox^{*}, David E. Graham[§], Ross Overbeek[‡], Marjory A. Snead^{*}, Martin Keller^{*}, Monette Aujay^{*}, Robert Huber^{||}, Robert A. Feldman^{*}, Jay M. Short^{*}, Gary J. Olsen[§] & Ronald V. Swanson^{*}

^{*} Diversa Corporation, 10665 Sorrento Valley Road, San Diego, California 92121, USA

[‡] Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

[§] Department of Microbiology, University of Illinois, Urbana, Illinois 61801, USA

^{||} Lehrstuhl für Mikrobiologie, Universität Regensburg W-8400, Regensburg W-8400, Germany

***Aquifex aeolicus* was one of the earliest diverging, and is one of the most thermophilic, bacteria known. It can grow on hydrogen, oxygen, carbon dioxide, and mineral salts. The complex metabolic machinery needed for *A. aeolicus* to function as a chemolithoautotroph (an organism which uses an inorganic carbon source for biosynthesis and an inorganic chemical energy source) is encoded within a genome that is only one-third the size of the *E. coli* genome. Metabolic flexibility seems to be reduced as a result of the limited genome size. The use of oxygen (albeit at very low concentrations) as an electron acceptor is allowed by the presence of a complex respiratory apparatus. Although this organism grows at 95 °C, the extreme thermal limit of the Bacteria, only a few specific indications of thermophily are apparent from the genome. Here we describe the complete genome sequence of 1,551,335 base pairs of this evolutionarily and physiologically interesting organism.**

Complete genome sequences have been determined for a number of organisms, including Archaea¹, Bacteria^{2–7}, and Eukarya⁸. Here we present and explore the genome sequence of *Aquifex aeolicus*. With growth-temperature maxima near 95 °C, *Aquifex pyrophilus* and *A. aeolicus* are the most thermophilic bacteria known. Although isolated and described only recently⁹, these species are related to filamentous bacteria first observed at the turn of the century, growing at 89 °C in the outflow of hot springs in Yellowstone National Park^{10,11}. The observation of these macroscopic assemblages would later be instrumental in the drive to culture hyperthermophilic organisms¹².

The *Aquificaceae* represent the most deeply branching family within the bacterial domain on the basis of phylogenetic analysis of 16S ribosomal RNA sequences^{13,14}, although analyses of individual protein sequences vary in their placement of *Aquifex* relative to other groups^{15–18}. The genera in this group, *Aquifex* and *Hydrogenobacter*, are thermophilic, hydrogen-oxidizing, microaerophilic, obligate chemolithoautotrophs^{9,19–21}. *A. aeolicus* (isolated by R.H. and K. O. Stetter) was cultured at 85 °C under an H₂/CO₂/O₂ (79.5:19.5:1.0) atmosphere in a medium containing only inorganic components. *A. aeolicus* does not grow on a number of organic substrates, including sugars, amino acids, yeast extract or meat extract. Unlike its close relative *A. pyrophilus*, *A. aeolicus* has not been shown to grow anaerobically with nitrate as an electron acceptor in the laboratory.

From study of the physiology of the organism, several predictions can be made. As an autotroph, *A. aeolicus* must have genes encoding proteins for one or more modes of carbon fixation and a complete set of biosynthetic genes. As autotrophy is a feature that is distributed throughout the Archaea and Bacteria, most of the associated genes are expected to be of ancient origin and clearly related to those characterized elsewhere. The obligate autotrophy suggests a biosynthetic rather than a degradative character. Oxygen respiration

implies the presence of corresponding utilization and tolerance genes. The early divergence of the *Aquificaceae* inferred from ribosomal RNA sequences leads to several questions. Are the machineries for oxygen usage and tolerance homologous to those found in mitochondria and well studied organisms such as *Escherichia coli*, or were they invented separately? If there was far less oxygen when the lineage originated, is there evidence for use of alternative oxidants?

Genome

General features of the *A. aeolicus* genome are listed in Box 1. We classified 1,512 open-reading frames (ORFs) into one of three categories, namely, identified (Table 1), hypothetical, or unknown. Identified ORFs were further classified into one of 57 cellular role categories adapted from Riley²² (Table 1). The relatively high G + C content of the two 16S-23S-5S rRNA operons (65%) is characteristic of thermophilic bacterial rRNAs²³. The genome is densely packed: most genes are apparently expressed in polycistronic operons and many convergently transcribed genes overlap slightly. Nonetheless, many genes that are functionally grouped within operons in other organisms, such as the tryptophan or histidine biosynthesis pathways, are found dispersed throughout the *A. aeolicus* genome or appear in novel operons. Even when they encode subunits of the same enzyme, the genes are often separated on the chromosome (for example, *gltB* and *gltD*, the genes encoding the large and small subunits of glutamate synthase). Operon organization of genes for the biosynthesis of amino acids is found in both Archaea and Bacteria but it is not universal in either group. *A. aeolicus* is extreme in that no two amino acid biosynthetic genes are found in the same operon. In contrast, genes required for electron transport, hydrogenase subunits, transport systems, ribosomal subunits, and flagella are often in functionally related operons in *A. aeolicus* (Fig. 1). No introns or inteins (protein splicing elements) were detected in the genome.

A single extrachromosomal element (ECE) was identified during sequencing. Sequence redundancy for the total project was calculated to be 4.83. The ECE, however, is significantly over-represented

[†] Present addresses: Codex Bioinformatics Services, PO Box 90273, San Diego, California 92169, USA (G.D.); Department of Bioinformatics, SmithKline Beecham Pharmaceuticals, Collegeville, Philadelphia 19426, USA (P.V.W.)

relative to the chromosome; when calculated independently for the final assemblies, redundancies are 4.73 and 8.76 for the chromosome and for the ECE, respectively. The ECE therefore appears to be present at roughly twice the copy number of the chromosome. Although no ORFs on the ECE can be assigned a function with confidence, except for a transposase, two of the predicted proteins show similarity to hypothetical proteins in the *Methanococcus jannaschii* genome¹. One ORF on the ECE is also present in two identical copies on the *A. aeolicus* chromosome, providing evidence of genetic exchange between the chromosome and the ECE.

Reductive tricarboxylic acid cycle

As an autotroph, *A. aeolicus* obtains all necessary carbon by fixing CO₂ from the environment. An assay for activity of the reductive tricarboxylic acid (TCA) cycle in *A. pyrophilus* cell extracts showed *in vitro* activities for each proposed reaction²⁴. The reductive (reverse) TCA cycle fixes two molecules of CO₂ to form acetyl-coenzyme A (acetyl-CoA) and other biosynthetic intermediates²⁵. The *A. aeolicus* genome contains genes encoding malate dehydrogenase, fumarate hydratase, fumarate reductase, succinate-CoA ligase, ferredoxin oxidoreductase, isocitrate dehydrogenase, aconitase and citrate synthase, which together could constitute the TCA pathway. There is no biochemical evidence for alternative carbon-fixation pathways in *A. pyrophilus*^{24,25} nor is there sequence evidence for such pathways in *A. aeolicus*.

The TCA cycle is vital as it provides the substrates of many biosynthetic pathways. (It is beyond the scope of this report to detail these biosynthetic pathways, but they seem to be typically bacterial¹, and candidate genes for all or most of the enzymes have been identified in *A. aeolicus*.) The central role of the TCA cycle is emphasized by duplication of many of its constituent genes in *A. aeolicus*. Two genes encode proteins that are similar to malate dehydrogenase (in addition to a lactate dehydrogenase). The fumarate hydratase is split into amino- and carboxy-terminal subunits, as is the case in *M. jannaschii*¹. Unlinked genes encoding two iron-sulphur proteins of fumarate reductase (alternatively succinate dehydrogenase) accompany a single flavoprotein subunit. Two sets of genes resembling succinate-CoA ligase (both the α - and β -subunits) are present. *A. aeolicus* has two putative operons encoding four-subunit (α , β , γ , δ) 2-acid ferredoxin oxidoreductases; members of this family catalyze reversible carboxylation/decarboxylation of pyruvate, 2-isoketoglutarate, or 2-oxoglutarate with varying specificity²⁶. These duplicated genes may encode paralogous proteins with unique substrate specificity, as opposed to redundant functions. For example, a paralogue of succinate-CoA ligase may activate citrate with coenzyme A to form citryl-CoA, which citrate synthase can cleave to produce oxaloacetate and acetyl-CoA.

Gluconeogenesis through the Embden–Meyerhof–Parnas pathway

Growing autotrophically, *A. aeolicus* must synthesize pentose and hexose monosaccharides from products of the reductive TCA cycle. Pyruvate produced by pyruvate ferredoxin oxidoreductase or by pyruvate carboxylase (oxaloacetate decarboxylase)²⁴ may enter the Embden–Meyerhof–Parnas pathway of glycolysis and gluconeogenesis. Genes encoding fructose-1,6-bisphosphatase, an essential gluconeogenic enzyme in *E. coli*, have not been identified in the genomes of the autotrophs *A. aeolicus* or *M. jannaschii*¹, suggesting that an unidentified pathway may exist. The *A. aeolicus* genome also encodes enzymes of the pentose-phosphate pathway and enzymes for glycogen synthesis and catabolism. We found neither (phospho) gluconate dehydrase nor 2-keto-3-deoxy-(6-phospho)gluconate aldolase of the Entner–Doudoroff pathway.

Respiration

Aquifex species are able to grow by using oxygen concentrations as low as 7.5 p.p.m. (R.H. and K. O. Stetter, unpublished observations).

The enzymes for oxygen respiration are similar to those of other bacteria: ubiquinol cytochrome *c* oxidoreductase (*bc*₁ complex), cytochrome *c* (three different genes) and cytochrome *c* oxidase (with two different subunit I genes and two different subunit II genes). The alternative system, with cytochrome *bd* ubiquinol oxidase, is also present. Clearly, the *Aquifex* lineage did not independently invent oxygen respiration. This leaves at least three possibilities: consistent with the ability of *Aquifex* to use very low levels of oxygen, the oxygen-respiration system was highly developed when oxygen had only a small fraction of its present concentration before the advent of oxygenic photosynthesis; contrary to what is implied by the 16S phylogeny, the lineage including *Aquifex* originated after the rise in atmospheric oxygen; or oxygen respiration developed once, and was then laterally transferred among bacterial lineages and acquired by *Aquifex*.

Many other oxidoreductases are present in addition to those obviously involved in oxygen respiration. The physiological role of most of these oxidoreductases is unknown or ambiguous, but two deserve comment. There is a putative nitrate reductase in the genome, although *A. aeolicus* has not been observed to perform NO₃⁻ respiration, unlike the closely related *A. pyrophilus*. The nitrate reductase gene is adjacent to a nitrate transporter, and may be involved in nitrogen assimilation rather than respiration. It is also possible that *A. aeolicus* has a latent ability to respire with nitrate but that the conditions required have not been found. Two gene sequences show strong similarities to Rieske proteins, even though the rest of the ubiquinol cytochrome *c* oxidoreductase subunits appear only once in the genome. One of these Rieske protein genes is adjacent to a sulphide dehydrogenase subunit, suggesting a role in sulphur respiration.

Oxidative stress

A. aeolicus grows optimally under microaerophilic conditions and consequently possesses various protective enzymes to counter reactive oxygen species, particularly superoxide and peroxide. The genome contains three genes encoding superoxide dismutases, two of the copper/zinc family and one of the iron/manganese family. The latter has also been noted in *A. pyrophilus*²⁷. One of the copper/zinc superoxide dismutase genes is located in a large gene cluster encoding formate dehydrogenase.

No catalase genes were identified. There are several genes in the genome that might encode proteins that catalyze the detoxification of H₂O₂, including cytochrome *c* peroxidase, thiol peroxidase, and two alkyl hydroperoxide reductase genes. All of these enzymes require an exogenous reductant and therefore do not evolve O₂. However, treatment of *A. pyrophilus*⁹ or *A. aeolicus* biomass with H₂O₂ results in the rapid evolution of gas bubbles. This catalase activity may result from a novel enzyme that cannot yet be identified by sequence similarity.

Motility

Like *A. pyrophilus*⁹, *A. aeolicus* is motile and possesses monopolar polytrichous flagella. More than 25 genes encoding proteins involved in flagellar structure and biosynthesis have been identified in *A. aeolicus* (Box 1). However, no homologues of the bacterial chemotaxis system were identified. In enteric bacteria, membrane-bound receptors bind chemoattractants and repellents and mod-

Figure 1 Linear map of the *A. aeolicus* circular chromosome. Genes are shown as arrows which denote the direction of transcription and are coloured to denote functional categorization according to the key below the figure. The sequences of the two rRNA gene clusters are identical. Here, the first base of the coding sequence of *fusA* was arbitrarily assigned as base number 1 as no origin of replication has been identified. ORF numbers are discontinuous because some ORFs representing 100 amino acids or more are not predicted to be coding and are not shown.

ulate the activity of the histidine kinase CheA²⁸. Phosphoryl groups from CheA are transferred to CheY, which then binds to the flagellar switch, altering the direction of flagellar rotation. Homologous chemotaxis systems are present in the archaea *Halobacterium salinarum*²⁹ and *Pyrococcus* sp. OT3 (H. Sizuya, personal communication), although the bacterial and archaeal flagellar apparatuses are not homologous³⁰. The *M. jannaschii* genome also lacks homologues of known genes required for chemotaxis. Thus, either motility in *A. aeolicus* and *M. jannaschii* is undirected or input for controlling taxis is mediated through another, unidentified system. The most studied chemotaxis systems respond to sugars and amino acids, although responses to other inputs (for example, metals, redox potential, and light) may also occur. In contrast to all the organisms known to possess the classical chemotactic signal-transduction pathways, both *A. aeolicus* and *M. jannaschii* are obligate chemoautotrophs. Chemoautotrophs may respond to a different set of factors, such as concentrations of dissolved gas (CO₂, H₂ or O₂) or another critical parameter such as temperature.

In *E. coli*, the flagellar switch is essential for flagellar structure and function and coupling of chemotaxis signals. But the *A. aeolicus* genome encodes homologues of only two of the three *E. coli* proteins that make up the switch, FliG and FliN. Biochemical³¹ and genetic³² studies implicate the missing FliM protein as the receptor for phosphorylated CheY, the switch signal. The absence of both FliM and CheY in *A. aeolicus* supports the identification of FliM as the receptor for phosphorylated CheY in *E. coli*. This result also argues against a direct role for FliM in torque generation.

DNA replication and repair

The *A. aeolicus* primary replicative DNA polymerase, corresponding to the DNA polymerase III holoenzyme in *E. coli*, probably consists

Figure 2 Histogram representation of the similarity of selected classes of predicted proteins to predicted proteins from the *E. coli* (EC) and *M. jannaschii* (MJ) genomes. Predicted *A. aeolicus* proteins representing each category were independently compared to sets of all potential polypeptides (≥ 100 amino acids) from the two genomes using FASTA⁴⁴. If the top scoring alignment covered $\geq 80\%$ of the length of the *A. aeolicus* protein, the score was plotted. There were more positives found in the *E. coli* genome in nearly every category. Hypothetical proteins (those identified by database match but of unknown function) are very similarly represented by *M. jannaschii* and *E. coli*. There are a small number of very highly conserved hypotheticals that are shared between *A. aeolicus* and *M. jannaschii*. Generally, biosynthetic categories show less discrimination than information-processing categories, which are clearly more *E. coli*-like. The variation in the apparent rates of evolution in different categories suggests that different phylogenies may be inferred depending on the sequence analysed. Within each graph, correspondence to *E. coli* is shown in white and *M. jannaschii* is shown in black. Avg id, average identity; count, number of proteins analysed.

Box 1 *Aquifex aeolicus* genome features

General

Length 1,551,335 bp
G + C content 43.4%
Protein-coding regions 93%
Stable RNA 0.8%
Non-coding repeats (none significant)
Intergenic sequences 6.2%

RNA

Ribosomal RNA Chromosome coordinates
16S-23S-5S 572785-567770
16S-23S-5S 1192069-1197084
Transfer RNA
44 species (7 clusters, 28 single genes)
Other RNAs Chromosome coordinates
tmRNA 1153844-1153498

Chromosomal coding sequences

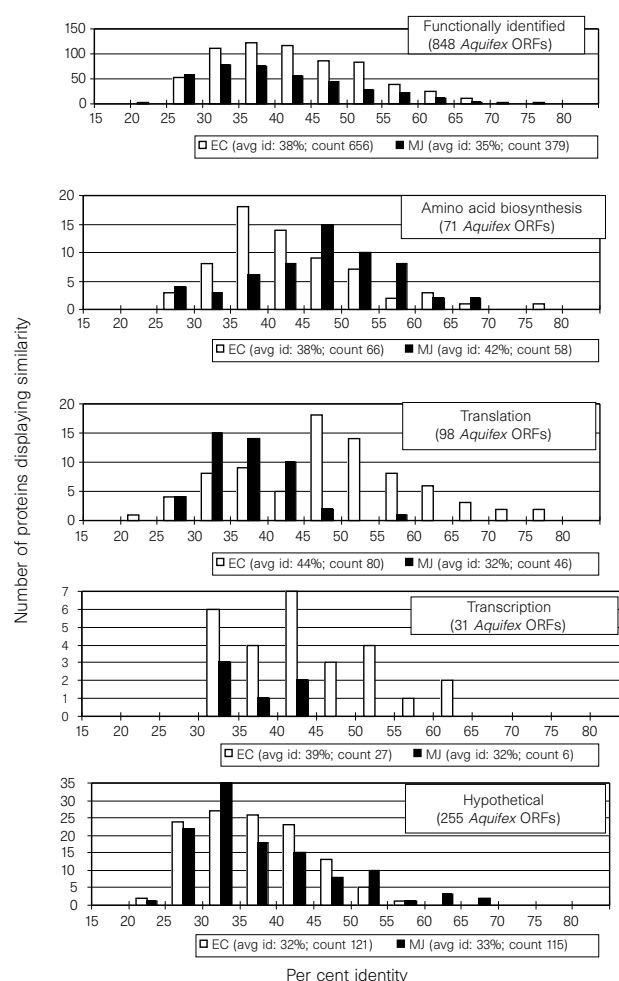
849 similar to protein of known function (average length 1,066 bp)
256 similar to protein of unknown function (average length 898 bp)
407 unknown coding regions (average length 762 bp)
1,512 total (average length 956)

Extrachromosomal element (ECE)

Length 39,456 bp
G + C content 36.4%
Protein-coding regions 53.5%

ECE-coding sequences

1 similar to proteins of known function (length 948 bp)
4 similar to proteins of unknown function (average length 667 bp)
27 unknown coding regions (average length 648 bp)



of a core structure containing α - and ϵ -subunits, a γ - τ -subunit and an additional member of the γ - τ / δ' -family. A gene encoding a protein homologous to the β -sliding clamp was also found. This minimalistic complex lacks homologous θ -, δ -, χ - and ψ -subunits, as does the *Mycoplasma genitalium* holoenzyme³. Translation of the 54K (relative molecular mass) γ - τ -ATPase subunit may proceed without a programmed frameshift to produce a protein similar to the N-terminal region of the *E. coli* γ -subunit. DNA polymerase I is present as separate Klenow fragment and 5' $\rightarrow$ 3' exonuclease subunits, encoded by two non-adjacent ORFs. Although the repair polymerase, DNA polymerase II, has not been found in *A. aeolicus*, one ORF (Aq1422) encodes a protein similar to the eukaryotic DNA repair polymerase- β . A member of the same family has been identified in *Thermus aquaticus*³³ and *Bacillus subtilis*.

Transcriptional and translational apparatuses

The transcriptional apparatus of *A. aeolicus* is similar to that of *E. coli* and lacks any components specific to the Eukarya or Archaea (Fig. 2). In addition to the core RNA polymerase α -, β -, and β' -subunits, four σ -factors which determine promoter specificity are present (Table 1). Several different families of bacterial transcriptional regulators were also identified, including two-component systems. All of the ribosomal proteins and elongation factors common to other bacteria are present, indicating that all bacteria-specific ribosomal proteins were present in the common ancestor of *Aquifex* and other bacteria. Also present are the four *sel* genes required for the cotranslational incorporation of selenocysteine. These latter genes are clustered in a 15-kilobase-pair segment that also encodes the biosynthetic and structural proteins for formate dehydrogenase, the only selenocysteine-containing protein identified. The gene that encodes selenocysteine transfer RNA, *selC*, is apparently cotranscribed with the genes encoding the formate dehydrogenase structural proteins.

A. aeolicus lacks glutamyl-tRNA and asparaginyl-tRNA synthetases. The genes required for transamidation of glutamyl-tRNA^{Gln} are present³⁴. Charging of asparaginyl-tRNA is likely to proceed through the analogous reaction, as shown in halobacteria³⁵, although the gene(s) for that transamidase are unknown. The canonical methionyl- and leucyl-tRNA synthetases have only been seen previously as single polypeptide enzymes; however, in *A.*

aeolicus the homologues appear fragmented into two subunits. In both cases, the genes that encode the N- and C-terminal portions are widely separated on the chromosome. No complete three-dimensional structural data are available for either methionyl- or leucyl-aminoacyl tRNA synthetases, but the subunit organization in the *A. aeolicus* aminoacyl-tRNA synthetases may reflect domain organization in the homologous proteins.

Thermophily

The *A. aeolicus* genome is the second completely sequenced genome of a hyperthermophile. By comparing the *A. aeolicus* and *M. jannaschii* genomes and contrasting them with the complete genomes of mesophiles, we can discover whether there are aspects of the genome or the encoded information that are diagnostic of hyperthermophiles. The G + C content of the stable RNAs is clearly indicative of the high growth temperature of the organism. This property can be used to identify stable RNAs against the relatively low G + C background of the *A. aeolicus* genome. The gene encoding tmRNA (or 10Sa RNA)³⁶, an RNA involved in tagging polypeptides translated from incomplete messenger RNAs for degradation, was located in this way.

Two genes for reverse gyrase are present in the genome. This is the only protein known to be present only in thermophiles. Other proteins, currently described as hypotheticals, may be diagnostic of hyperthermophiles but the data sets are not yet large enough to decide this with confidence.

Although features of stabilization may not be apparent in any given protein³⁷, a large enough data set may reveal general trends in amino-acid usage that are informative. Particularly important in this regard is inclusion of multiple genomes of hyperthermophiles so as not to allow the idiosyncracies of a single organism to bias the conclusions. As shown in Table 2, comparison of the amino-acid composition encoded by six genomes shows that use of individual amino acids can vary significantly from genome to genome. The data suggest trends that may be correlated with the thermostability of the encoded proteins. One apparent trend is that the hyperthermophile genomes encode higher levels of charged amino acids on average than mesophile genomes³⁸, primarily at the expense of uncharged polar residues. Glutamine in particular seems to be significantly discriminated against in the hyperthermophiles. Although this observation might be rationalized on the basis of

Table 2 Comparison of relative amino acid compositions (in percentages) of mesophiles and thermophiles

Amino acid	Mesophiles				Thermophiles	
	<i>H. influenzae</i>	<i>H. pylori</i>	<i>E. coli</i>	<i>Synechosystis</i>	<i>A. aeolicus</i>	<i>M. jannaschii</i>
A	8.21	6.83	9.55	9.07	5.90	5.54
C	1.03	1.09	1.11	1.01	0.79	1.27
D	4.98	4.77	5.20	5.07	4.32	5.52
E	6.48	6.88	5.91	6.20	9.63	8.67
F	4.46	5.41	3.87	3.75	5.13	4.20
G	6.65	5.76	7.42	7.77	6.75	6.41
H	2.05	2.12	2.26	1.93	1.54	1.43
I	7.10	7.20	5.95	6.31	7.32	10.45
K	6.32	8.94	4.48	4.26	9.40	10.36
L	10.50	11.18	10.56	10.93	10.57	9.38
M	2.44	2.28	2.86	2.12	1.92	2.33
N	4.89	5.83	3.88	3.76	3.60	5.24
P	3.72	3.28	4.41	5.09	4.07	3.38
Q	4.64	3.70	4.42	5.26	2.04	1.44
R	4.47	3.46	5.58	5.18	4.91	3.85
S	5.84	6.81	5.67	5.46	4.79	4.46
T	5.20	4.37	5.35	5.53	4.21	4.06
V	6.68	5.59	7.11	7.10	7.93	6.85
W	1.12	0.70	1.48	1.30	0.93	0.71
Y	3.12	3.68	2.83	2.78	4.13	4.33
Mesophiles				Thermophiles		
Charged residues (DEKRH)	24.11				29.84	
Polar/uncharged residues (GSTNQYC)	31.15				26.79	
Hydrophobic residues (LMIVWPAF)	44.74				43.36	

an increased rate of deamidation of this residue at higher temperatures, asparagine does not appear subject to similar discrimination.

Phylogeny

The placement of the *Aquifex* lineage as one of the earliest divergences in the eubacterial tree^{13,14} is interesting because of the insights it could provide into the ancestral eubacterial phenotype, including the hypothesized thermophilic nature of the first bacteria. Protein-based phylogenies often do not support the original rRNA-based placement^{15,16,18}. Thus, the availability of some 1,500 genes from an *Aquifex* species would seem to offer a definitive resolution of the phylogeny. However, our analyses of ribosomal proteins, aminoacyl-tRNA synthetases, and other proteins do not do so, showing no consistent picture of the organism's phylogeny. We cannot make a more complete analysis and discussion here, but some observations can be made. These proteins do not yield a statistically significant placement of the *Aquifex* lineage or of other major eubacterial lineages. This situation partially reflects the inadequacy of some protein sequences as indicators of distant molecular genealogy because of their particular evolutionary dynamic, including the patterns and rates of amino-acid replacements. In some cases (such as the aminoacyl-tRNA synthetases for arginine, cysteine, histidine, proline and tyrosine), the analyses are further complicated by the presence of paralogous genes and/or apparent lateral gene transfers. It seems that a more extensive survey of genes and a better sampling of major eubacterial taxa will be required to confidently confirm or refute an early divergence of the *Aquifex* lineage.

Conclusions

Advances in sequencing techniques have allowed us to move beyond studies of single genes to studies of complete genomes only recently². This rapid advance has created the opportunity to begin to characterize an organism with the full knowledge of the genome in hand. The complete genome summarized in this report represents our first view of *A. aeolicus*. The challenge now is to ask specific questions in ways which take advantage of the whole-genome data.

Beyond studies of any single organism in isolation, complete genomes allow comprehensive comparisons between organisms. For instance, comparisons of the similarity of genes can be made that reveal that genes in different categories vary in their relative conservation (Fig. 2). In addition, genome-wide trends are apparent. For example, why is there not more of a tendency to group functionally related genes (for example, biosynthetic pathways) into operons in *A. aeolicus*? This was also seen in the genome sequence of the autotroph *M. jannaschii*¹. Is this because the autotrophic lifestyle decreases the need for selective regulation? There also seem to be a few multifunctional, fused proteins in *A. aeolicus* and *M. jannaschii*. Although this seems unlikely to be related to autotrophy, it might be associated with extreme thermophily. The large number of diverse genome sequences that will become available in the coming years will allow more detailed correlation of global genomic properties with particular physiologies. □

Methods

Sequencing strategy. The sequencing strategy used to assemble the complete genome was based on the whole genome random (or 'shotgun') approach, which has been successfully used for other genomes of similar size¹⁻⁴. Shotgun sequencing projects are characterized by two phases: an initial completely random phase in which the bulk of the data is collected, followed by a closure phase where directed techniques are used to close gaps and complete the assembly. By pursuing a strategy where only 97% coverage was initially achieved, we were able to limit the number of sequences needed for the random phase to only 10,500 (ref. 39).

Sequences were generated from a small insert library constructed in λ ZAP II vectors^{40,41} (average insert length 2.9 kilobase pairs). Two different methods were used for sequencing: first, dye-primer M13-21 and M13 reverse primer ABI Prism CS⁺ ready reaction kits, analysed on 48-cm 4% polyacrylamide

gels; and second, dye-terminator (ABI Prism FS⁺) reactions using two pBluescript-specific primers. These reactions were analysed on 36-cm 5% Long-Ranger gels.

The sequence fragments were assembled on an Apple Power Macintosh computer using Sequencher (Gene Codes, Ann Arbor, MI), an assembly and editing program. Assembly was typically performed in batches of roughly 200–400 sequences, and was followed by inspection and editing of the assemblies. All sequences in the set were compared with all others through this process. After assembly, the sequences comprised ~750 contigs at the end of the random phase. Sequences were obtained from both ends of ~200 randomly chosen clones from a fosmid library^{42,43}. These sequences were then assembled with consensus sequences derived from the contigs of random-phase sequences using Sequencher. Gaps between contigs were closed by direct sequencing on fosmids not wholly contained within a contig. The fosmid library thus served a purpose analogous to that of the λ -scaffold in other projects¹⁻⁴. The final eight gaps were closed by direct sequencing of polymerase chain reaction (PCR) products generated with the TaqPlus Long PCR System (Stratagene Cloning Systems, La Jolla, CA).

Consequences of reducing the number of sequences in the random phase are the large number of gaps that remain to be closed in the directed phase, and the reduction in overall coverage. To ensure that reduced coverage did not compromise accuracy, ~200 oligonucleotide primers were synthesized to resequence regions of ambiguity identified by visual inspection of the entire assembly. 13,785 sequences, with an average edited read length of 557 base pairs, constitute the final assembly. On the basis of a relatively small number of errors identified during the annotation process, we estimate the error frequency to be <0.01%, comparable to other published genomic sequence estimates.

Gene (ORF + RNA) identification and functional assignment approaches.

Coding regions of the *A. aeolicus* genome were analysed and assigned using primarily the programs BLASTP⁴⁴ and FASTA⁴⁵ to search against a non-redundant protein database. Many analyses were carried out within the context of MAGPIE^{46,47}, an integrated computing environment for genome analysis. The results of these analyses are available for user interpretation, validation, and categorization. Additional ORFs were identified and start sites refined using the program CRITICA (J. H. Badger and G.J.O., unpublished program). Finally, all presumed 'intergenic regions' were examined with BLASTX for similarities to known protein sequences⁴⁸. Transfer RNA genes were identified with the program tRNAscan-SE⁴⁹.

Received 26 August 1997; accepted 3 February 1998.

1. Bult, C. et al. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073 (1996).
2. Fleischmann, R. D. et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–511 (1995).
3. Fraser, C. M. et al. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403 (1995).
4. Tomb, J.-F. et al. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547 (1997).
5. Himmelreich, R. et al. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24**, 4420–4449 (1996).
6. Kaneko, T. et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC7803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136 (1996).
7. Blattner, F. R. et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).
8. Goffeau, A. et al. Life with 6000 genes. *Science* **274**, 546 (1996).
9. Huber, R. et al. *Aquifex pyrophilus* gen. nov. sp. nov. represents a novel group of marine hyperthermophilic hydrogen oxidizing bacteria. *Arch. Microbiol.* **15**, 340–351 (1992).
10. Reysenbach, L., Wickham, G. S. & Pace, N. R. Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Appl. Environ. Microbiol.* **60**, 2113–2119 (1994).
11. Setchell, W. A. The upper temperature limits of life. *Science* **17**, 934–937 (1903).
12. Brock, T. D. The road to Yellowstone—and beyond. *Annu. Rev. Microbiol.* **49**, 1–28 (1995).
13. Burggraf, S., Olsen, G. J., Stetter, K. O. & Woese, C. R. A phylogenetic analysis of *Aquifex pyrophilus*. *Syst. Appl. Microbiol.* **15**, 353–356 (1992).
14. Pitulle, C. et al. Phylogenetic position of the genus *Hydrogenobacter*. *Int. J. Syst. Bacteriol.* **44**, 620–626 (1994).
15. Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc. Natl Acad. Sci. USA* **93**, 7749–7754 (1996).
16. Klenk, H.-P., Palm, P. & Zillig, W. in *Molecular Biology of the Archaea* (eds Pfeifer, F., Palm, P. & Schleifer, K. H.) 139–147 (Vch Pub, 1994).
17. Bocchetta, M. et al. Arrangement and nucleotide sequence of the gene (*fus*) encoding elongation factor G (EF-G) from the hyperthermophilic bacterium *Aquifex pyrophilus*: phylogenetic depth of hyperthermophilic bacteria inferred from analysis of the EF-G/*fus* sequences. *J. Mol. Evol.* **41**, 803–812 (1995).
18. Wetmur, J. G. et al. Cloning, sequencing, and expression of RecA proteins from three distantly related thermophilic eubacteria. *J. Biol. Chem.* **269**, 25928–25935 (1994).
19. Kawasumi, T., Igarashi, Y., Kodama, T. & Minoda, Y. *Hydrogenobacter thermophilus* gen. nov., sp. nov.

- an extremely thermophilic, aerobic, hydrogen-oxidizing bacterium. *Int. J. Syst. Bacteriol.* **34**, 5–10 (1984).
20. Kristjansson, J., Ingason, A. & Alfredsson, G. A. Isolation of thermophilic obligately autotrophic hydrogen-oxidizing bacteria, similar to *Hydrogenobacter thermophilus*, from Icelandic hot springs. *Arch. Microbiol.* **140**, 321–325 (1985).
21. Kryukov, V. R., Savel'eva, N. D. & Pusheva, M. A. *Calderobacterium hydrogenophilum* gen. nov., sp. nov. an extreme thermophilic bacterium and its hydrogenase activity. *Microbiology (Engl. Trans. Mikrobiologiya)* **52**, 611–618 (1983).
22. Riley, M. Functions of the gene products of *Escherichia coli*. *Microbiol. Rev.* **57**, 862–952 (1993).
23. Weisburg, W. G., Giovannoni, S. J. & Woese, C. R. The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. *Syst. Appl. Microbiol.* **11**, 128–134 (1989).
24. Beh, M., Strauss, G., Huber, R., Stetter, K. O. & Fuchs, G. Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaeobacterium *Thermoproteus neutrophilus*. *Arch. Microbiol.* **160**, 306–311 (1993).
25. Fuchs, G. in *Autotrophic Bacteria* (eds Schegel, H. G. & Bowein, B.) 365–382 (Springer, New York, 1987).
26. Mai, X. & Adams, M. W. Characterization of a fourth type of 2-keto acid-oxidizing enzyme from a hyperthermophilic archaeon: 2-ketoglutarate ferredoxin oxidoreductase from *Thermococcus litoralis*. *J. Bacteriol.* **178**, 5890–5896 (1996).
27. Lim, J. H. *et al.* Cloning and expression of superoxide dismutase from *Aquifex pyrophilus*, a hyperthermophilic bacterium. *FEBS Lett.* **406**, 142–146 (1997).
28. Bourret, R. B., Borkovich, K. A. & Simon, M. I. Signal transduction pathways involving protein phosphorylation in prokaryotes. *Annu. Rev. Biochem.* **60**, 401–441 (1991).
29. Rudolph, J., Tolliday, N., Schmitt, C., Schuster, S. C. & Oesterhelt, D. Phosphorylation in halobacterial signal transduction. *EMBO J.* **14**, 4249–4257 (1995).
30. Jarrell, K. F., Bayley, D. P. & Kostyukova, A. S. The archaeal flagellum: a unique motility structure. *J. Bacteriol.* **178**, 5057–5064 (1996).
31. Welch, M., Oosawa, K., Aizawa, S. I. & Eisenbach, M. Effects of phosphorylation, Mg^{2+} , and conformation of the chemotaxis protein CheY on its binding to the flagellar switch protein FlIM. *Biochemistry* **33**, 10470–10467 (1994).
32. Sockett, H., Yamaguchi, S., Kihara, M., Irikura, V. M. & Macnab, R. M. Molecular analysis of the flagellar switch protein FlIM of *Salmonella typhimurium*. *J. Bacteriol.* **174**, 793–806 (1992).
33. Motoshima, H. *et al.* Molecular cloning and nucleotide sequence of the aminopeptidase T gene of *Thermus aquaticus* YT-1 and its high-level expression in *Escherichia coli*. *Agric. Biol. Chem.* **54**, 2385–2392 (1990).
34. Curnow, A. W. *et al.* Glu-tRNA_{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl Acad. Sci. USA* **94**, 11819–11826 (1997).
35. Curnow, A. W., Ibba, M. & Söll, D. tRNA-dependent asparagine formation. *Nature* **382**, 589–590 (1996).
36. Tu, G. F., Reid, G. E., Zhang, J. G., Moritz, R. L. & Simpson, R. J. C-terminal extension of truncated proteins in *Escherichia coli* with a 10Sa decapeptide. *J. Biol. Chem.* **270**, 9322–9326 (1995).
37. Böhm, G. & Jaenicke, R. Relevance of sequence statistics for the properties of extremophilic proteins. *Int. J. Pept. Protein Res.* **43**, 97–106 (1994).
38. Choi, I.-G. *et al.* Random sequence analysis of genomic DNA of a hyperthermophile: *Aquifex pyrophilus*. *Extremophiles* **1**, 125–134 (1997).
39. Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239 (1988).
40. Short, J. M., Fernandez, J. M., Sorge, J. A. & Huse, W. D. Lambda ZAP: a bacteriophage lambda expression vector with *in vivo* excision properties. *Nucleic Acids Res.* **16**, 7583–7600 (1988).
41. Alting-Mees, M. A. & Short, J. M. pBluescript II: gene mapping vectors. *Nucleic Acids Res.* **17**, 9494 (1989).
42. Shizuya, H. *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA* **89**, 8794–8797 (1992).
43. Kim, U.-J., Shizuya, H., de Jong, P. J., Birren, B. & Simon, M. I. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res.* **20**, 1083–1085 (1992).
44. Altschul, S. F., Fish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
45. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).
46. Gaasterland, T. & Sensen, C. W. MAGPIE: automated genome interpretation. *Trends Genet.* **12**, 76–78 (1996).
47. Gaasterland, T. & Sensen, C. W. Fully automated genome analysis that reflects user needs and preferences. A detailed introduction to the MAGPIE system architecture. *Biochimie* **78**, 302–310 (1996).
48. Gish, W. & States, D. J. Identification of protein coding regions by database similarity search. *Nature Genet.* **3**, 266–272 (1993).
49. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).

Acknowledgements. This work was supported in part by Department of Energy Microbial Genome Program grants (to R.V.S., C. R. Woese and G.J.O.). We thank C. Woese for his cooperation in the analysis of the genome and interest in the project; K. Stetter for continuing interest; G. Frey, J. Holaska, S. Peralta, D. Hafenbrandl, S. Delk, T. Robinson, and J. Arnett for technical assistance; and D. Robertson, J. Stein, I. Sanyal, T. Richardson, G. Hauska, and K. Williams for discussions.

Correspondence should be addressed to R.V.S. (e-mail: rswanson@diversa.com). Requests for *Aquifex aeolicus* should be addressed to R.H. (e-mail: Robert.huber@biologie.uni-regensburg.de). The sequences have been deposited with GenBank and assigned accession numbers AE000657 (chromosome) and AE000667 (extrachromosomal element).



Table 1 *Aquifex aeolicus* Open Reading Frame Identifications. Gene numbers (Aq) correspond to those in Fig.1. Percentages refer to the identity found in the best FASTA alignment. The percentage of the sequence covered by the alignment is displayed with bullets as follows 20–40% • • , 40–60% • • • , 60–80% • • • • , 80–100% • • • • •

Amino Acid Biosynthesis																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			</
-------------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

Nature © Macmillan Publishers Ltd 1998

Aq1708	pkA	phosphofructokinase	49.4%	Aq046	pyrD	dihydroorotate dehydrogenase	50.5%
Aq750	pgi	glucose-6-phosphate isomerase	37.8%	Aq1305	pyrDB	dihydroorotate dehydrogenase electron transfer subunit	34.7%
Aq118	pgk	phosphoglycerate kinase	54.5%	Aq1580	pyrF	orotidine-5'-phosphate decarboxylase	37.2%
Aq1990	pgmA	phosphoglycerate mutase	33.2%	Aq1334	pyrG	GTP synthetase	57.5%
Aq501	pnu	phosphoglucosyltransferase/phosphomannomutase	37.8%	Aq1713	pyrH	UMP kinase	62.1%
Aq2142	ppsA	phosphoenolpyruvate synthase	56.3%	Aq640	thy	thymidylate synthase complementing protein	30.5%
Aq1520	pycA	pyruvate carboxylase c-terminal domain	46.6%	Aq969	tmk	thymidylate kinase	35.1%
Aq1517	pycB	pyruvate carboxylase n-terminal domain	57.1%	Aq1907	umpS	uridine 5-monophosphate synthase	42.1%
Aq360	timA	triose phosphate isomerase	52.2%	Aq2163	uraP	uracil phosphoribosyltransferase	42.0%
Hydrogenase				Regulation			
Aq665	hoxZ	Ni/Fe hydrogenase B-type cytochrome subunit	40.4%	Aq1058	acrR1	transcriptional regulator (TetR/AcrR family)	34.1% ..
Aq667	hupD	HupD hydrogenase related function	40.9%	Aq2179	acrR2	transcriptional regulator (TetR/AcrR family)	31.0%
Aq666	hupE	HupE hydrogenase related function	38.3%	Aq281	acrR3	transcriptional regulator (TetR/AcrR family)	29.7%
Aq1021	hypA	hydrogenase accessory protein HypA	39.8%	Aq1387	arsR	transcriptional regulator (ArsR family)	35.3%
Aq671	hypB	hydrogenase expression/formation protein B	50.6%	Aq1724	degT	transcriptional regulator (DegT/Dnr/EryC family)	34.1%
Aq1157	hypD	hydrogenase expression/formation protein HypD	56.1%	Aq534	draG	ADP-ribosylglycohydrolase	32.1%
Aq662	mbhL1	hydrogenase large subunit	50.6%	Aq831	exsB	trans-regulatory protein ExsB	38.5%
Aq960	mbhL2	hydrogenase large subunit	44.3%	Aq490	fmr	transcriptional regulator (Crp/Fnr family)	29.5%
Aq804	mbhL3	hydrogenase large subunit	27.9%	Aq1207	furR1	transcriptional regulator (FurR family)	37.9%
Aq660	mbhS1	hydrogenase small subunit	66.6%	Aq1418	furR2	transcriptional regulator (FurR family)	34.6%
Aq965	mbhS2	hydrogenase small subunit	51.3%	Aq213	ghnBi	PH-like protein GlnBi	48.0% ..
Aq802	mbhS3	hydrogenase small subunit	36.7%	Aq1908	hix	GTP-binding protein HIX	40.3%
Aq1591	shyS	soluble hydrogenase small subunit	41.6%	Aq1115	hksP1	histidine kinase sensor protein	27.7% ..
Sugar metabolism				Aq316	hksP2	histidine kinase sensor protein	28.1%
Aq968	cbfE2	ribulose-5-phosphate 3-epimerase	47.2%	Aq905	hksP3	histidine kinase sensor protein	23.6%
Aq1658	fucA1	fuculose-1-phosphate aldolase	31.8%	Aq231	hksP4	histidine kinase sensor protein	28.2%
Aq1979	fucA2	fuculose-1-phosphate aldolase	29.7%	Aq1156	hoxX	hydrogenase regulation HoxX	46.7%
Aq498	gnd	6-phosphogluconate dehydrogenase	45.2%	Aq093	hth	transcriptional regulator (H-T-H)	50.2%
Aq497	gsdA	glucose-6-phosphate 1-dehydrogenase	32.3%	Aq1019	hypE	hydrogenase expression/formation protein	44.3%
Aq1138	rpiB	ribose 5-phosphate isomerase B	54.5%	Aq672	hypF	transcriptional regulatory protein HypF	44.8%
Aq119	talC	transaldolase	71.1%	Aq764	icR	transcriptional regulator (IcR family)	30.4%
Aq1765	tktA	transketolase	52.4%	Aq638	lysR1	transcriptional regulator (LysR family)	32.8%
NADH dehydrogenase				Aq1038	lysR2	transcriptional regulator (LysR family)	28.9%
Aq1385	nuoA1	NADH dehydrogenase I chain A	42.0%	Aq702	merR	transcriptional regulator (MerR family)	32.8%
Aq1310	nuoA2	NADH dehydrogenase I chain A	44.9%	Aq218	niifA	transcriptional regulator (NiifA family)	42.8%
Aq1312	nuoB	NADH dehydrogenase I chain B	60.1% ..	Aq1117	ntrC1	transcriptional regulator (NtrC family)	41.0%
Aq551	nuoD1	NADH dehydrogenase I chain D	37.7%	Aq1792	ntrC2	transcriptional regulator (NtrC family)	40.2%
Aq1314	nuoD2	NADH dehydrogenase I chain D	42.2%	ntrC3	ntrC3	transcriptional regulator (NtrC family)	40.0%
Aq574	nuoE	NADH dehydrogenase I chain E	36.8%	Aq164	ntrC4	transcriptional regulator (NtrC family)	38.3%
Aq573	nuoF	NADH dehydrogenase I chain F	20.5% ..	Aq2069	obg	GTP-binding protein	54.9% ..
Aq437	nuoG	NADH dehydrogenase I chain G	35.4% ..	Aq319	phoB	transcriptional regulator (PhoB-like)	41.6%
Aq1315	nuoH1	NADH dehydrogenase I chain H	41.0%	Aq906	phoU	transcriptional regulator (PhoU-like)	41.9%
Aq1373	nuoH2	NADH dehydrogenase I chain H	42.1%	Aq844	spoT	(p)ppGpp 3-pyrophosphohydrolase	57.2%
Aq1374	nuoH3	NADH dehydrogenase I chain H	38.9%	Aq1496	xyIR	transcriptional regulator (NagC/XyIR family)	29.3%
Aq1317	nuoI	NADH dehydrogenase I chain I	30.5%	DNA Replication and Repair			
Aq1375	nuoI2	NADH dehydrogenase I chain I	29.2%	Aq358	dinG	ATP-dependent helicase (DinG family)	27.9%
Aq1318	nuoJ1	NADH dehydrogenase I chain J	35.4%	Aq322	dnaA	chromosome replication initiator protein DnaA	36.5%
Aq1377	nuoJ2	NADH dehydrogenase I chain J	30.6%	Aq1472	dnaB	replicative DNA helicase	40.3%
Aq1319	nuoK1	NADH dehydrogenase I chain K	51.1%	Aq910	dnaC	DNA replication protein DnaC	26.4%
Aq1378	nuoK2	NADH dehydrogenase I chain K	48.4%	Aq1008	dnaE	DNA polymerase III alpha subunit	41.9%
Aq1320	nuoL1	NADH dehydrogenase I chain L	39.0%	Aq1493	dnaG	DNA primase	39.8%
Aq866	nuoL2	NADH dehydrogenase I chain L	30.2%	Aq1882	dnaN	DNA polymerase III beta chain	32.1%
Aq1379	nuoL3	NADH dehydrogenase I chain L	43.1%	dnaQ	dnaQ	DNA polymerase III epsilon subunit	40.0%
Aq1321	nuoM1	NADH dehydrogenase I chain M	43.6%	Aq1855	dnaX	DNA polymerase III gamma subunit	36.6%
Aq1382	nuoM2	NADH dehydrogenase I chain M	36.9%	Aq1422	dpbF	DNA polymerase beta family	39.1%
Aq1322	nuoN1	NADH dehydrogenase I chain N	34.1%	Aq1693	dpfF	N-terminus of factor SPO1 DNA polymerase	37.3%
Aq1383	nuoN2	NADH dehydrogenase I chain N	32.8%	Aq980	gyrA	DNA gyrase A subunit	43.6%
Lipid metabolism				Aq1026	gyrB	gyrase B	55.2% ..
Aq2058	aas	2-acylglycerophosphoethanolamine acyltransferase	37.1%	Aq2057	helX	DNA helicase	49.7%
Aq1206	accA	acetyl-CoA carboxylase alpha subunit	57.1%	Aq1484a	himA	DNA binding protein HU	40.2%
Aq1363	accB	biotin carboxyl carrier protein	44.6%	Aq2174	ihfB	integration host factor beta subunit	35.8%
Aq1664	accC1	biotin carboxylase	54.4%	Aq1394	lig	DNA ligase (ATP dependent)	50.8%
Aq1470	accC2	biotin carboxylase	56.5%	Aq633	ligA	DNA ligase (NAD dependent)	45.7%
Aq445	accD	acetyl-CoA carboxyltransferase beta subunit	56.9%	Aq1578	mutL	DNA mismatch repair protein MutL	72.3%
Aq1717a	acpP	acyl carrier protein	71.2%	Aq308	mutS1	DNA mismatch repair protein MutS	77.5%
Aq813	acpS	holo-[acyl-carrier protein] synthase	30.8%	Aq1242	mutS2	DNA mismatch repair protein MutS	37.0%
Aq2104	acs	acetyl-coenzyme A synthetase	54.0%	Aq1449	mutT	8-Oxo-dGTPase domain (mutT domain)	46.3%
Aq2103	acs'	acetyl-coenzyme A synthetase c-terminal fragment	61.2%	Aq282	mutY1	endonuclease III	53.6%
Aq1249	cds	phosphatidate cytidylyltransferase	29.2%	Aq172	mutY2	endonuclease III	51.8%
Aq1737	cfa	cyclopropane-fatty-acyl-phospholipid synthase	37.5%	Aq496	mutY3	endonuclease III	43.4%
Aq892	fahD	malonyl-CoA:acyl carrier protein transacylase	42.1%	Aq1629	nfo	deoxyribonuclease IV	39.0%
Aq1712	fabF	3-oxoacyl-[acyl-carrier-protein] synthase II	58.4%	Aq1495	ngt	thermoococcal nuclease homolog	36.4%
Aq1716	fabG	3-oxoacyl-[acyl-carrier-protein] reductase	52.9%	Aq1628	pol	DNA polymerase I 3'-5' exo domain	43.2%
Aq1099	fabH	3-oxoacyl-[acyl-carrier-protein] synthase III	47.0%	Aq1967	polA	DNA polymerase I (PolI)	30.5%
Aq1552	fabI	enoyl-[acyl-carrier-protein] reductase (NADH)	49.6%	Aq1610	radC	DNA repair protein RadC	39.0%
Aq056	fabZ	(3R)-hydroxymyristoyl-acyl carrier protein dehydratase	58.7%	Aq2150	recA	recombination protein RecA	88.5%
Aq999	fadD	long-chain-fatty-acid CoA ligase	30.0%	Aq2053	recG	ATP-dependent DNA helicase RecG	38.9%
Aq1638	lplA	lipase-protein lipase A	28.1% ..	Aq2155	recJ	single-strand-DNA-specific exonuclease RecJ	31.8%
Aq958	pgsA	phosphatidylglycerophosphate synthase	37.3% ..	Aq561	recN	recombination protein RecN	27.7%
Aq2154	pgsA	phosphatidylglycerophosphate synthase	38.9%	Aq1478	recR	recombination protein RecR	38.3%
Aq1101	plxK	PlxK protein	43.7%	Aq793	rep	ATP-dependent DNA helicase REP	33.4%
Purines, Pyrimidines, Nucleotides and Nucleosides				Aq1886	abcD	ATP-dependent dsDNA exonuclease	29.9%
Aq094	nrdA	ribonucleotide reductase alpha chain	35.0%	Aq064	ssb	single stranded DNA-binding protein	39.4%
Aq1505	nrdF	ribonucleotide reductase beta chain	36.2% ..	Aq657	topA	topoisomerase I	39.6%
Purines				Aq1159	topG1	reverse gyrase	41.6%
Aq568	deoD	purine nucleoside phosphorylase	33.1%	Aq886	topG2	reverse gyrase	35.1%
Aq236	guaA	GMP synthase	58.4%	Aq686	uvrA	repair excision nuclease subunit A	61.0%
Aq2023	guaB	inosine monophosphate dehydrogenase	65.4%	Aq1856	uvrB	repair excision nuclease subunit B	53.9%
Aq544	hpt	hypoxanthine-guanine phosphoribosyltransferase	48.2%	Aq2126	uvrC	repair excision nuclease subunit C	32.5%
Aq078	katA	adenylate kinase	50.0%	Transcription			
Aq1590	ndk	nucleoside diphosphate kinase	48.2%	RNA polymerase and transcription factors			
Aq1636	prs	phosphoribosylpyrophosphate synthetase	55.2%	Aq613	deaD	ATP-dependent RNA helicase DeaD	42.3%
Aq1290	purA	adenylosuccinate synthetase	49.2%	Aq357a	flgM	anti sigma factor FlgM	20.6%
Aq597	purB	adenylosuccinate lyase	52.4%	Aq1218	flaA	RNA polymerase sigma factor FlaA	37.2%
Aq2117	purC	phosphoribosylaminimidazole-succinocarboxamide synthase	52.5%	Aq259	nusA	transcription termination NusA	45.4%
Aq742	purD	phosphoribosylamine-glycine ligase	54.2%	Aq133	nusB	transcription termination NusB	32.3%
Aq1178	purE	phosphoribosylaminimidazole carboxylase	64.6%	Aq1931	nusG	transcription antitermination protein NusG	46.3% ..
Aq1175	purF	amidophosphoribosyltransferase	42.7%	Aq873	rho	transcriptional terminator Rho	59.6%
Aq1963	purH	phosphoribosylaminimidazolecarboxamide formyltransferase	48.2%	Aq070	rpoA	RNA polymerase alpha subunit	40.4%
Aq245	purK	phosphoribosyl aminoimidazole carboxylase	35.6%	Aq1939	rpoB	RNA polymerase beta subunit	44.0%
Aq1836	purL	phosphoribosylformylglycinamide synthase II	49.3%	Aq1945	rpoC	RNA polymerase beta prime subunit	46.9%
Aq769	purM	phosphoribosylformylglycinamide cyclo-ligase	50.0%	Aq1490	rpoD	RNA polymerase sigma factor RpoD	41.6%
Aq857	purN	phosphoribosylformylglycinamide formyltransferase	48.3%	Aq599	rpoN	RNA polymerase sigma factor RpoN	30.6%
Aq1105	purQ	phosphoribosyl formylglycinamide synthase I	51.1%	Aq1452	rpoS	RNA polymerase sigma factor RpoS	40.5%
Aq1818	purU	formyltetrahydrofolate deformylase	56.3%	RNA modification			
Pyrimidines				Aq1816	kagA	dimethyladenosine transferase	36.1%
Aq410	carA	carbamoyl phosphate synthetase small subunit	52.2%	Aq1067	miaA	RNA delta-2-isopentenylpyrophosphate (IPP) transferase	38.2%
Aq1172	carB	carbamoyl-phosphate synthase large subunit	60.7%	Aq411	pcnB1	poly A polymerase	28.5%
Aq2101	carB	carbamoyl-phosphate synthase, large subunit	63.1%	Aq2158	pcnB2	poly A polymerase	33.9% ..
Aq2153	cmk	cytidylate kinase	38.5%	Aq221	phpA	polyribonucleotide nucleotidyltransferase	45.0%
Aq1607	dcd	deoxycytidine triphosphate deaminase	39.5%	Aq894	queA	queuosine biosynthesis protein	46.9%
Aq220	dut	deoxyuridine 5-triphosphate nucleotidohydrolase	39.5%	Aq946	rnc	RNase III	35.8%
Aq409	pyrB	aspartate carbamoyltransferase catalytic chain	42.0%	Aq1955	rnH8	RNase HII	48.4%
Aq806	pyrC	dihydroorotase	37.3%	Aq924	rnphI	RNase PH	64.0%
				Aq1661	spoU	rRNA methylase SpoU	44.0%
				Aq1308	tgt	guanine (RNA)-ribosyltransferase	52.6%
				Aq841	trm1	N2,N2-dimethylguanosine tRNA	

Aq1489	trmD	methyltransferase	34.6% ***	Aq1671	hslV	heat shock protein HslV	57.6% ***
Aq749	truA	tRNA guanine-N1 methyltransferase	42.9% ***	Aq1450	htrA	periplasmic serine protease	38.3% ***
Aq705	truB	pseudouridine synthase 1	33.1% ***	Aq242	lon	Lon protease	50.6% ***
Aq1890	tsrR	tRNA pseudouridine 55 synthase	38.2% ***	Aq076	map	methionyl aminopeptidase	44.1% ***
Aq2046	vacB	rRNA methylase	36.4% ***	Aq1459	npr	neutral protease	27.7% **
Aq257	ycgA	YacB protein (ribonuclease II family)	37.9% ***	Aq2099	pepA	leucine aminopeptidase	39.5% ***
		RNA methyltransferase (TrmA-family)	28.8% ***	Aq1535	pepQ	xaa-pro dipeptidase	31.9% ***
				Aq618	pipI	protease I	41.8% ***
Translation				Aq797	prc	carboxyl-terminal protease	41.8% ***
Aq2131	fnt	methionyl-tRNA formyltransferase	45.7% ***	Aq552	sms	ATP-dependent protease sms	46.2% ***
Aq247	gatA	glutamyl-tRNA(Gln) amidotransferase subunit A	53.6% ***	Aq2204	ymxG	processing protease	28.3% ***
Aq461	gatB	glutamyl-tRNA(Gln) amidotransferase subunit B	48.8% ***				
Aq2147a	gatC	glutamyl-tRNA(Gln) amidotransferase subunit C	41.1% ***	Transport			
Aq346	pth	peptidyl-tRNA hydrolase	48.8% ***	Aq122	abcT1	ABC transporter	34.7% ***
Aminoacyl tRNA synthetases				Aq620	abcT2	ABC transporter	36.8% ***
Aq1293	alaS	alanyl-tRNA synthetase	46.6% ***	Aq1095	abcT3	ABC transporter (ABC-C subfamily)	34.4% ***
Aq923	argS	arginyl-tRNA synthetase	39.4% ***	Aq1094	abcT4	ABC transporter	37.7% ***
Aq1677	aspS	aspartyl-tRNA synthetase	51.3% ***	Aq1097	abcT5	ABC transporter (hylB subfamily)	5.5% ***
Aq1068	cysS	cysteinyl-tRNA synthetase	45.0% ***	Aq417	abcT6	ABC transporter	51.8% ***
Aq763	genX	lysyl-tRNA synthetase (genX) homolog	38.6% ***	Aq413	abcT7	ABC transporter	51.5% ***
Aq1221	glxX	glutamyl-tRNA synthetase	48.5% ***	Aq297	abcT8	ABC transporter	49.3% ***
Aq945	glyQ	glycyl-tRNA synthetase alpha subunit	61.9% ***	Aq2160	abcT9	ABC transporter	45.3% ***
Aq2141	glyS	glycyl-tRNA synthetase beta subunit	37.1% ***	Aq1531	abcT10	ABC transporter	36.4% ***
Aq1225	hisS1	histidyl-tRNA synthetase	43.3% ***	Aq2122	abcT11	ABC transporter	42.5% ***
Aq1155	hisS2	histidyl-tRNA synthetase	34.9% ***	Aq2137	abcT12	ABC transporter	38.2% ***
Aq305	ileS	isoleucyl-tRNA synthetase	82.1% ***	Aq1563	abcT13	ABC transporter (MsbA subfamily)	30.5% ***
Aq351	leuS	leucyl-tRNA synthetase alpha subunit	50.7% ***	Aq695	acrD1	cation efflux system (AcrB/AcrD/AcrF family)	22.7% ***
Aq1770	leuS2	leucyl-tRNA synthetase beta subunit	47.2% ***	Aq1122	acrD2	cation efflux system (AcrB/AcrD/AcrF family)	32.0% ***
Aq2102	lysU	lysyl-tRNA synthetase	53.2% ***	Aq469	acrD3	cation efflux system (AcrB/AcrD/AcrF family)	34.2% ***
Aq1257	metG	methionyl-tRNA synthetase alpha subunit	45.0% ***	Aq786	acrD4	cation efflux (AcrB/AcrD/AcrF family)	27.7% ***
Aq422	metG2	methionyl-tRNA synthetase beta subunit	64.2% ***	Aq112	amtB	ammonium transporter	49.0% ***
Aq953	pheS	phenylalanyl-tRNA synthetase alpha subunit	51.9% ***	Aq682	arsA1	anion transporting ATPase	41.5% ***
Aq1730	pheS2	phenylalanyl-tRNA synthetase beta subunit	35.4% ***	Aq343	arsA2	anion transporting ATPase	33.9% ***
Aq365	proT	proline-tRNA synthetase	44.1% ***	Aq851	craX	(Zg ²⁺) and Czc(2 ⁺) transport protein	31.1% ***
Aq298	serS	seryl-tRNA synthetase	59.4% ***	Aq724	craI	cation transporting ATPase (E1-E2 family)	30.7% ***
Aq1667	thrS	threonyl-tRNA synthetase	48.5% ***	Aq1445	craA2	cation transporting ATPase (E1-E2 family)	28.1% ***
Aq992	trpS	tryptophanyl-tRNA synthetase	38.4% ***	Aq1125	craA3	cation transporting ATPase (E1-E2 family)	43.8% ***
Aq1751	tyrS	tyrosyl-tRNA synthetase	56.2% ***	Aq1132	czcB1	cation efflux system (czcB-like)	23.7% ***
Aq1413	valS	valyl-tRNA synthetase	33.2% ***	Aq1331	czcB2	cation efflux system (czcB-like)	26.9% ***
				Aq468	czcB2	cation efflux system (czcB-like)	28.5% ***
				Aq1073	czcD	cation efflux system (CzcD-like)	43.4% ***
Ribosomal Proteins				Aq911	ebS	erythrocyte band 7 homolog	50.2% ***
Aq1935	rplA	ribosomal protein L01	57.9% ***	Aq1062	emrB	major facilitator family transporter	28.3% ***
Aq013	rplB	ribosomal protein L02	46.9% ***	Aq1255	feoB	ferrous iron transport protein	32.6% ***
Aq009	rplC	ribosomal protein L03	51.3% ***	Aq1330	glpP	proton/sodium-glutamate symport protein	35.6% ***
Aq011	rplD	ribosomal protein L04	33.8% ***	Aq1268	hwsT	high affinity sulfate transporter	29.4% ***
Aq1652	rplE	ribosomal protein L05	67.0% ***	Aq1863	kch	potassium channel protein	30.1% ***
Aq1649	rplF	ribosomal protein L06	46.2% ***	Aq1725	lepA	G-protein LepA	59.8% ***
Aq2042	rplH	ribosomal protein L07	36.5% ***	Aq1229	mifT	transporter (major facilitator family)	37.2% ***
Aq1936	rplK	ribosomal protein L10	71.4% ***	Aq447	mgfC	Mg(2+) transport ATPase	36.2% ***
Aq1933	rplL	ribosomal protein L11	75.4% ***	Aq1609	modA	molybdenum periplasmic binding protein	38.2% ***
Aq1937	rplM	ribosomal protein L12/L12	60.6% ***	Aq086	modC	Molybdenum transport system permease	44.8% ***
Aq1877	rplN	ribosomal protein L13	59.5% ***	Aq415	napA1	Na(+)/H(+) antiporter	27.6% ***
Aq1654	rplO	ribosomal protein L14	57.4% ***	Aq929	napA2	Na(+)/H(+) antiporter	32.7% ***
Aq018	rplP	ribosomal protein L16	59.3% ***	Aq2030	napA3	Na(+)/H(+) antiporter	26.8% ***
Aq069	rplQ	ribosomal protein L17	48.7% ***	Aq215	nasA	nitrate transporter	35.8% ***
Aq1648	rplR	ribosomal protein L18	62.7% ***	Aq1441	oppA	transporter (extracellular solute binding protein family 5)	37.0% ***
Aq1954	rplS	ribosomal protein L19	59.8% ***				
Aq952	rplT	ribosomal protein L20	63.5% ***	Aq481	oppB	transporter (OppB/C family)	46.2% ***
Aq016a	rplV	ribosomal protein L22	47.3% ***	Aq1509	oppC	oligopeptide transport system permease	46.2% ***
Aq012	rplW	ribosomal protein L23	52.2% ***	Aq2019	pstA	phosphate transport system permease PstA	43.5% ***
Aq1653	rplX	ribosomal protein L24	50.8% ***	Aq1055	pstB	phosphate transport ATP binding protein	68.1% ***
Aq1644	rpmD	ribosomal protein L30	46.4% **	Aq2018	pstC	phosphate transport system permease protein C	45.2% ***
Aq1930a	rpmG	ribosomal protein L33	57.9% ***	Aq2016	pstS	phosphate-binding periplasmic protein	52.4% ***
Aq792a	rpmI	ribosomal protein L35	48.3% ***	Aq2129	shf	Na(+)-dependent transporter (Shf family)	34.9% ***
Aq1485	rpsA	ribosomal protein S01	32.6% ***	Aq098	secG	protein export membrane protein SecG	35.7% ***
Aq2007	rpsB	ribosomal protein S02	60.3% ***	Aq2077	snf	Na(+)-neurotransmitter symporter (Snf family)	25.7% ***
Aq017	rpsC	ribosomal protein S03	54.0% ***	Aq2106	snf	Na(+)-solute symporter (Snf family)	47.4% **
Aq072	rpsD	ribosomal protein S04	51.9% ***	Aq1988	tolQ	TolQ homolog	32.5% **
Aq1645	rpsE	ribosomal protein S05	60.6% ***	Aq1504	trkI	K+ transport protein homolog	40.6% ***
Aq063	rpsF	ribosomal protein S06	32.7% ***	Aq031	trnS	transporter (Pho87 family)	46.8% ***
Aq1832	rpsG1	ribosomal protein S07	52.5% ***	Uncategorized			
Aq734	rpsG2	ribosomal protein S07	51.9% ***	Aq1023	acuC1	acetoin utilization protein	36.9% ***
Aq1651	rpsH	ribosomal protein S08	39.9% ***	Aq2110	acuC2	acetoin utilization protein	38.6% ***
Aq1878	rpsI	ribosomal protein S09	50.5% ***	Aq158	apfA	APfA hydrolase	36.6% ***
Aq008	rpsJ	ribosomal protein S10	55.9% ***	Aq458	bcp	bacterioferritin comigratory protein	40.6% ***
Aq073	rpsK	ribosomal protein S11	60.7% ***	Aq542	bcpC	phosphonopyruvate decarboxylase	37.4% ***
Aq735	rpsL1	ribosomal protein S12	78.9% ***	Aq147	cobW	cobalamin synthesis related protein CobW	29.5% ***
Aq1834	rpsL2	ribosomal protein S12	78.9% ***	Aq1303a	cspC	cold shock protein	67.2% ***
Aq074	rpsM	ribosomal protein S13	61.9% ***	Aq1265	cstA	carbon starvation protein A	33.0% ***
Aq1651a	rpsN	ribosomal protein S14	51.6% ***	Aq348	cic	general stress protein Cic	34.7% ***
Aq226a	rpsO	ribosomal protein S15	61.6% ***	Aq212	cysS	cyanate hydrolase	39.5% ***
Aq123	rpsP	ribosomal protein S16	36.6% ***	Aq337	cysQ	CysQ protein	47.4% ***
Aq020	rpsQ	ribosomal protein S17	59.6% ***	Aq528	dedF	phenylacetyl acid decarboxylase	52.4% ***
Aq064a	rpsR	ribosomal protein S18	48.5% ***	Aq148	devC	deoxyribose-phosphate aldolase	46.6% ***
Aq015	rpsS	ribosomal protein S19	63.1% **	Aq2095	dksA	dnaK suppressor protein	35.1% **
Aq1767	rpsT	ribosomal protein S20	40.0% ***	Aq1994	eraI	GTP-binding protein EraI	49.7% ***
Aq867a	rpsU	ribosomal protein S21	38.2% ***	Aq1919	eraE	GTP binding protein Era	43.0% ***
Translation factors				Aq1540	gcpE	GcpE protein	50.1% ***
Aq1364	efp	elongation factor P	48.6% ***	Aq1052	gcsH1	glycine cleavage system protein H	28.6% ***
Aq2114	eif	initiation factor eIF-2B alpha subunit	58.4% ***	Aq1657	gcsH2	glycine cleavage system protein H	39.8% ***
Aq712	frr	ribosome recycling factor	43.0% ***	Aq944	gcsH3	glycine cleavage system protein H	36.7% ***
Aq001	fusA	elongation factor EF-G	91.9% ***	Aq1108	gcsH4	glycine cleavage system protein H	44.8% ***
Aq075a	infA	initiation factor IF-1	69.1% ***	Aq1458	gcvT	aminomethyltransferase	42.3% **
Aq002	infB	initiation factor IF-2	68.2% ***				

The seeds of rich galaxy clusters in the Universe

F. Governato*, C. M. Baugh*, C. S. Frenk*, S. Cole*,
C. G. Lacey†, T. Quinn‡ & J. Stadel‡

* Physics Department, Durham University, Durham DH1 3LE, UK

† Theoretical Astrophysics Centre, Copenhagen, Denmark

‡ Astronomy Department, University of Washington, Seattle, Washington 98195, USA

The discovery¹ of a population of young galaxies at a redshift when the Universe was about a tenth of its current age has shed new light on the question of when and how galaxies formed. Within the context of popular models², this is the population of primeval galaxies that built themselves up to the size of present-day galaxies through the process of repeated mergers called hierarchical clustering. But the recent detection³ of a large concentration of these primeval galaxies appears to be incompatible with hierarchical clustering models, which generally predict that clusters of this size are fully formed later in time. Here we use a combination of theoretical techniques—semi-analytic modelling and *n*-body simulations—to show that such large concentrations should be quite common in a universe dominated by cold dark matter, and that they are the progenitors of the rich galaxy clusters seen today. We predict the clustering properties of primeval galaxies which should, when compared with data that will be collected in the near future, test our current understanding of galaxy formation within the framework of a universe dominated by cold dark matter.

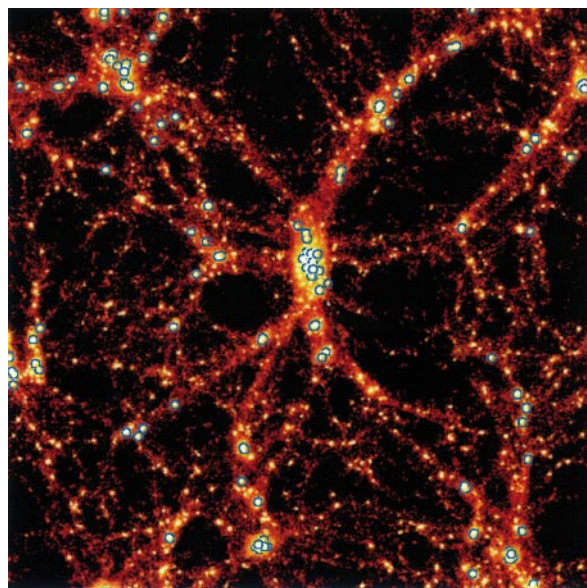
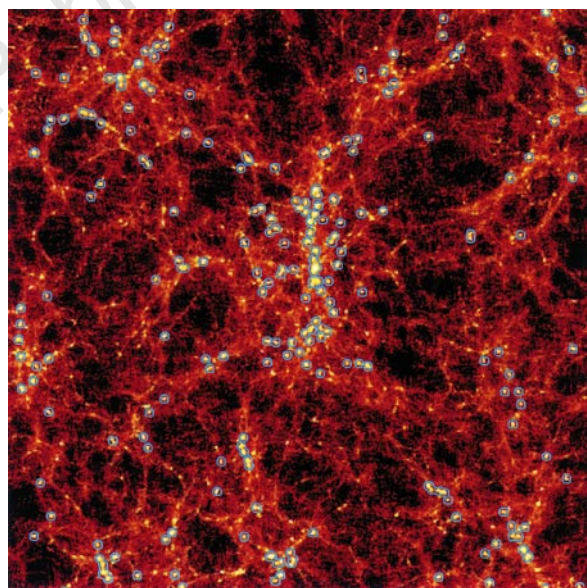
The long-standing search for primeval galaxies was recently brought to fruition by the combination of deep imaging and Keck Telescope spectroscopy⁴. Because they were originally identified by a sharp break in their spectrum, at the wavelength corresponding to the limit of the Lyman series of hydrogen, these galaxies are called 'Lyman break' galaxies. The most distant examples have redshift $z \approx 3.5$, corresponding to the time when the Universe was only about 10% of its current age. The structure discovered by Steidel and collaborators³ appears as a spike in the redshift distribution at $z = 3.09$. It contains 15 Lyman-break galaxies in a 9×18 arcmin

field³. The co-moving spatial dimensions of this structure are huge, approximately $8 \times 10 \times 13 h^{-1}$ Mpc in a universe with critical density. (Throughout, we denote Hubble's constant as $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$.)

The current model for the formation of cosmic structure is the cold dark matter theory⁵. For suitable values of the parameters, this theory provides a good description of large-scale structure, from scales probed by microwave background anisotropy measurements to those mapped by galaxy surveys⁶. In this theory, galaxies form by gas cooling and condensing into dark-matter haloes which grow by accretion and mergers in a hierarchical fashion. There is growing observational evidence that young galaxies are indeed built up from several small pieces^{7,8}, as predicted by the theory.

The technique of semi-analytic modelling provides an effective means to describe and quantify the complex physics involved in galaxy formation and evolution. The processes that need to be considered are the shock heating and radiative cooling of gas within collapsing dark-matter haloes, the subsequent transformation of cold gas into stars, the regulation of star formation by feedback from stellar winds and supernovae, and the merging of galaxy fragments. All these processes can be encoded in a set of simple, physically

Figure 1 The formation and growth of large structures in a cold dark matter universe. Top panel, the distribution of mass and Lyman-break galaxies in a slice of an $\Omega = 1$ CDM simulation at $z = 3$. The slice has side $50 h^{-1}$ Mpc and thickness $5 h^{-1}$ Mpc. The region plotted contains a large cluster at $z = 0$. The simulation was performed using a parallel tree-code with 3 million particles. The high resolution allows dark-matter haloes as small as that of the Large Magellanic Cloud to be resolved. The initial amplitude of density fluctuations was chosen so as to reproduce approximately the observed abundance of rich galaxy clusters today². The logarithm of the dark-matter density is colour-coded so that the highest-density regions are white and the lowest-density regions are black. The blue circles mark the locations where, according to our semi-analytic galaxy formation model, Lyman-break galaxies have formed. Whenever an individual dark-matter halo contains more than one Lyman-break galaxy, the first one is placed at the centre of the halo and the rest at the positions of randomly selected halo particles. According to the semi-analytic model, there are 900 Lyman-break galaxies in the entire $(50 h^{-1} \text{ Mpc})^3$ simulation volume. As expected², both our $\Omega = 1$ and $\Omega_0 = 0.3$ models produce an abundance of Lyman-break galaxies at high redshift consistent with the observed counts. The Lyman-break galaxies are preferentially found in the higher-density regions and, as a result, they are strongly biased relative to the underlying mass distribution. Bottom panel, the same slice as in the top panel evolved to the present day. The blue circles mark the positions of particles which were associated with Lyman-break galaxies at $z = 3$. The largest halo in the simulation volume has a mass comparable to that of the Coma cluster and contains the descendants of 60 Lyman-break galaxies. The general appearance of the $\Omega = 0.3$ simulation is qualitatively similar.



motivated rules. A surprisingly small number of free parameters are involved, and these are set by reference to a small subset of local galaxy data. The result is a fully specified model of galaxy formation that can be used to make predictions for the properties of the high-redshift universe. Semi-analytic models have been remarkably successful in reproducing or predicting a range of galaxy properties, such as the luminosity function, number counts and colours^{9,10}, and the evolution of the Hubble sequence of morphological types^{11,12}. In particular, predictions for the cosmic history of star formation^{10,13}, were subsequently found to be in good agreement with observations¹⁴.

The semi-analytic approach on its own provides little direct information about the spatial distribution of galaxies. To overcome this limitation we applied the semi-analytic galaxy formation rules to dark-matter haloes grown in high resolution n -body simulations, an approach previously used to study the clustering of the local

galaxy population¹⁵. The supercomputer simulations that we analysed followed the formation of structure by gravitational instability in two cold dark matter (CDM) universes, one with the critical density, $\Omega = 1$, and the other with present-day density parameter $\Omega_0 = 0.3$ (and $h = 0.75$). These span the range of viable models of structure formation of this type. Dark-matter haloes were identified in the simulations at redshift $z = 3$ and populated with visible galaxies according to the predictions of the semi-analytic model for haloes of each mass². In this way we constructed catalogues of the positions, velocities and spectrophotometric properties of all the galaxies present in the simulations, including the effect on the colours of the galaxies produced by absorption by intervening gas¹⁶. The model galaxies were then 'observed' in the same filters used by Steidel and collaborators, and mock Lyman-break galaxies were identified in exactly the same way as in the real survey³.

The spatial distribution of the dark matter and Lyman-break

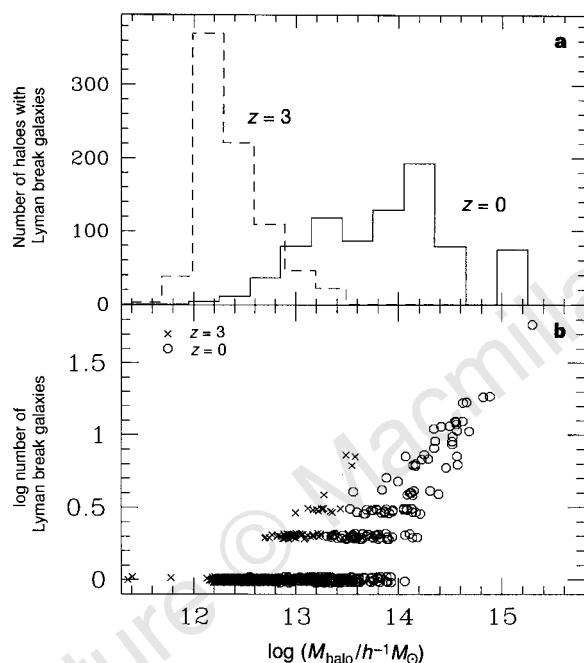


Figure 2 The mass distribution of the dark haloes that host Lyman break galaxies and their descendants. **a**, The mass distributions of dark-matter haloes that host Lyman-break galaxies at redshift 3 (dashed lines) and of haloes that host their descendants at the present time (solid line). At high redshift, Lyman-break galaxies are preferentially located within the most massive haloes that have formed at that time. These haloes tend to merge together to form the most massive structures present at $z = 0$. Today, the descendants of Lyman-break galaxies are to be found in a range of environments, from groups to clusters, with only a small fraction surviving in smaller, isolated haloes. **b**, The number of Lyman-break galaxies as a function of halo mass at $z = 3$ (crosses). Around 10% of the haloes that host Lyman-break galaxies contain more than one of these objects. The circles show the number of Lyman-break progenitors as a function of halo mass at $z = 0$. The most massive cluster in the simulation volume had 60 progenitors that were identified as Lyman-break galaxies at $z = 3$.

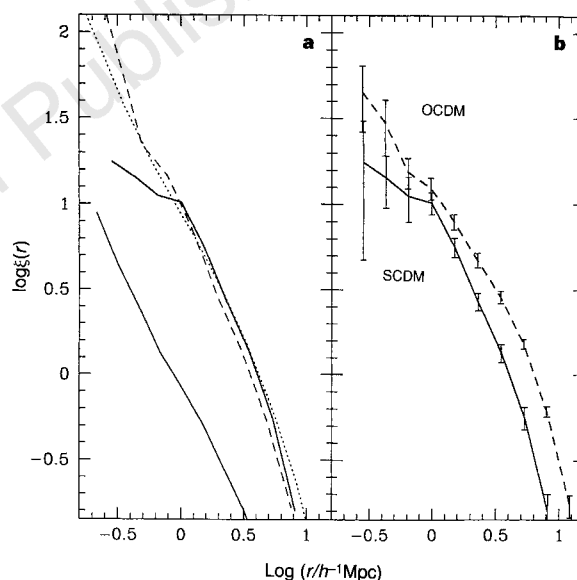


Figure 3 The clustering strength of Lyman break galaxies. **a**, Correlation functions, $\xi(r)$, in the $\Omega = 1$ CDM simulation at $z = 3$. The abscissa gives the pair separation in co-moving coordinates. The dashed line shows the true spatial correlation function of the Lyman-break galaxies. The thick solid line shows the apparent correlation function in redshift space—the correlations are reduced on small scales because of smearing by random motions within virialized haloes and are slightly enhanced on large scales owing to coherent inflow onto collapsing structures. The dotted line shows the analytic prediction² for the real-space correlation function. Overall, the agreement with the simulation result is very good. On small scales the simulation results are slightly higher, reflecting small inaccuracies in the analytic model. On large scales the simulation results are slightly suppressed by finite box effects in our relatively small simulation volume. On these large scales the analytic results are more reliable. The thin solid line shows the CDM correlation function at $z = 3$. **b**, The redshift space correlation functions in our $\Omega = 1$ (SCDM, solid line) and $\Omega_0 = 0.3$ (OCOM, dashed line) simulations. The error bars were obtained by bootstrap resampling. The co-moving clustering length of Lyman-break galaxies (defined as the pair separation at which $\xi(r) = 1$) at $z = 3$ is comparable to the clustering length of bright galaxies today. Perhaps counterintuitively, the Lyman-break galaxy correlation functions in the two models are quite similar, with characteristic clustering lengths which differ by only $\sim 50\%$. (The inclusion of a non-zero cosmological constant makes very little difference to the results of an $\Omega_0 \approx 0.3$ model; compare Fig. 8 of ref. 2. This is due in part to the normalization of the simulations which requires that they should approximately reproduce the observed abundance of clusters at the present day²⁷.

galaxies in our $\Omega = 1$ simulation is illustrated in Fig. 1. The top panel shows a slice through the simulation at $z = 3$, 1.6 Gyr after the Big Bang in this cosmology for $h = 0.5$. Lyman-break galaxies are seen to have formed within the largest dark-matter haloes present at that epoch and to trace the densest ridges of the dark-matter distribution. A redshift survey covering the central region of this slice would reveal a 'spike' like that observed by Steidel and collaborators³. The fate of the Lyman-break galaxies in this 'spike' is illustrated in the bottom panel of Fig. 1 which shows the same slice, but at the present day, 11.5 Gyr later. The material in the region of the spike has collapsed to form the large cluster seen near the centre of the image. This cluster has a mass of $9 \times 10^{14} h^{-1} M_{\odot}$, comparable to the mass of the Coma cluster¹⁷, and contains the descendants of 60 objects originally identified as Lyman-break galaxies.

The example in Fig. 1 illustrates the tendency for Lyman-break galaxies to form preferentially in large dark-matter haloes. The dashed line in Fig. 2a shows the predicted mass distribution of haloes that hosted Lyman-break galaxies at $z = 3$. The solid lines give the corresponding distribution for the present-day haloes in which the Lyman-break galaxies ended up. According to the theory, today's descendants of Lyman-break galaxies show a marked preference for dense environments, ranging from poor groups to rich clusters, with only a small fraction surviving as isolated galaxies. Thus, as conjectured by Steidel and co-workers³, large concentrations of Lyman-break galaxies at high redshift pick out regions where proto-groups and proto-clusters are forming. Our simulations clearly show that the precursors of clusters are highly irregular and elongated, reflecting their continuing formation by accretion of massive lumps at the intersection of long filaments. We note that the field observed by Steidel and collaborators³ contains a quasar, supporting the view that dense environments favour the formation of active galaxies¹⁸.

With our simulations we are able to calculate the probability of finding structures as large as the one discovered by Steidel and co-workers³. To do this, we first located galaxies in cylindrical 'skewers' of cross-sectional area equal to the area of the survey, using the simulation volumes and their periodic replications to sample a unit redshift interval. We then smoothed the galaxy density field with a top-hat filter of width $\Delta z = 0.04$, equal to the bin-width in the actual survey. Finally, we obtained the frequency of distinct peaks with 3.4 times the mean number of Lyman-break galaxies, which is the overdensity of the most significant spike in the real survey. We find that 45% of all skewers contain at least one such spike in a unit redshift interval centred on $z = 3$ and a significant number contain 2 or more. These numbers are very similar in the two CDM models we have examined. We conclude therefore that structures comparable to the spike recently discovered are not at all surprising in cold dark matter models irrespective of the detailed values of the cosmological parameters.

A more quantitative comparison between theory and observations is provided by the correlation function of Lyman-break galaxies, a statistic which is, in principle, measurable from the large surveys currently underway. Previous theoretical attempts to quantify clustering at high redshift have been restricted to dark-matter haloes^{19,20}. With the semi-analytic approach, on the other hand, we are able to predict directly the clustering properties of the observable galaxies themselves. There are substantial differences between the clustering statistics of dark haloes and Lyman-break galaxies due, for example, to scatter in the relation between halo mass and galaxy luminosity and to the fact that $\sim 10\%$ of dark haloes contain multiple Lyman-break galaxies (Fig. 2b). Our predicted correlation function for Lyman-break galaxies at redshift $z = 3$ is plotted in Fig. 3. Its amplitude is over 10 times larger than that of the underlying mass distribution, reflecting the preferential formation of Lyman-break galaxies in dark-matter haloes of mass $\geq 10^{12} h^{-1} M_{\odot}$ (which are rare at $z \approx 3$), and consistent with

analytic predictions². The bias parameter (defined as the ratio of r.m.s. fluctuations in the galaxies and mass respectively in spheres of radius $8h^{-1}$ Mpc) is $b = 4.2 \pm 0.4$ for the $\Omega = 1$ model and $b = 2.5 \pm 0.2$ in the $\Omega_0 = 0.3$ model.

Cold dark matter models of galaxy formation predict that only a small fraction of the stars present today—perhaps as little as 5%—have formed by $z \approx 3$ (ref. 10). In spite of this, the observed abundance of Lyman-break galaxies can be accounted for in these models², although the results depend sensitively on the amplitude of density fluctuations, the choice of stellar initial mass function and the possible effects of dust on galaxy luminosities. Thus, unless the effects of dust have been severely underestimated, the discovery of the Lyman-break galaxies indicates that the formation sites of most of the stars in our Universe have now been identified. According to theory, the Lyman-break galaxies are the brightest, most massive, examples of the population of young galaxies present at $z \approx 3$. Because of this, they are expected to be clustered much more strongly than the underlying dark matter, an expectation that, as we have shown, is consistent with the detection of a large spike in their redshift distribution. These data provide incontrovertible evidence for biased galaxy formation. At lower redshifts, other surveys²¹ generally pick out less extreme representatives of the galaxy population which tend to be less strongly biased than the Lyman-break galaxies are at $z \approx 3$. Using the theoretical approach that we have outlined here it is possible to relate data on galaxy clustering at various redshifts to theoretical expectations. Such comparisons, particularly with forthcoming surveys at high redshift, will test the model of galaxy formation in a cold dark matter universe. $\square$

Received 2 October; accepted 29 December 1997.

- Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M. & Adelberger, K. Spectroscopic confirmation of a population of normal galaxies at redshifts $z > 3$. *Astrophys. J.* **462**, L17–L22 (1996).
- Baugh, C. M., Cole, S., Frenk, C. S. & Lacey, C. G. The epoch of galaxy formation. *Astrophys. J.* (in the press).
- Steidel, C. C. et al. A large structure of galaxies at $z \sim 3$ and its cosmological implications. *Astrophys. J.* **492**, 428–438 (1998).
- Steidel, C. C., Pettini, M. & Hamilton, D. Lyman imaging of high-redshift galaxies. III. New observations of four QSO fields. *Astron. J.* **110**, 2519–2536 (1995).
- Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M. The evolution of large-scale structure in a universe dominated by cold dark matter. *Astrophys. J.* **292**, 371–394 (1985).
- Scott, D., Silk, J. & White, M. From microwave anisotropies to cosmology. *Science* **268**, 829–835 (1995).
- Pascarelle, S. M., Windhorst, R. A., Keel, W. C. & Odewahn, S. C. Sub-galactic clumps at a redshift of 2.39 and implications for galaxy formation. *Nature* **383**, 45–50 (1996).
- Lowenthal, J. D. et al. Keck spectroscopy for redshift $z \sim 3$ galaxies in the Hubble Deep Field. *Astrophys. J.* **481**, 673–688 (1997).
- Kauffmann, G., White, S. D. M. & Guiderdoni, B. The formation and evolution of galaxies within merging dark matter haloes. *Mon. Not. R. Astron. Soc.* **264**, 201–218 (1993).
- Cole, S., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F. & Zepf, S. E. A recipe for galaxy formation. *Mon. Not. R. Astron. Soc.* **271**, 781–806 (1994).
- Baugh, C. M., Cole, S. & Frenk, C. S. Evolution of the Hubble sequence in hierarchical models for galaxy formation. *Mon. Not. R. Astron. Soc.* **283**, 1361–1378 (1996).
- Kauffmann, G. The age of elliptical galaxies and bulges in a merger mode. *Mon. Not. R. Astron. Soc.* **281**, 487–492 (1996).
- White, S. D. M. & Frenk, C. S. Galaxy formation through hierarchical clustering. *Astrophys. J.* **379**, 52–79 (1991).
- Madau, P. et al. High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to $z \sim 4$. *Mon. Not. R. Astron. Soc.* **283**, 1388–1404 (1996).
- Kauffmann, G., Nusser, A. & Steinmetz, M. Galaxy formation and large-scale bias. *Mon. Not. R. Astron. Soc.* **268**, 795–811 (1997).
- Madau, P. Radiative transfer in a clumpy universe: The colors of high-redshift galaxies. *Astrophys. J.* **441**, 18–27 (1995).
- White, S. D. M., Navarro, J. F., Evrard, A. E. & Frenk, C. S. The baryon content of galaxy clusters—a challenge to cosmological orthodoxy. *Nature* **366**, 429–433 (1993).
- Ellington, E., Yee, H. K. C., Green, R. F. & Kinman, T. D. Clusters of galaxies associated with quasars. I—3C 206. *Astron. J.* **97**, 1539–1549 (1989).
- Mo, H. J. & Fukugita, M. Constraints on the cosmic structure formation models from early formation of giant galaxies. *Astrophys. J.* **467**, L9–L13 (1996).
- Bagla, J. S. Evolution of galaxy clustering. *Mon. Not. R. Astron. Soc.* (in the press).
- Lefevre, O. et al. The Canada–France redshift survey: VIII. Evolution of the clustering of galaxies from $z \approx 1$. *Astrophys. J.* **461**, 534–545 (1996).
- Elke, V. R., Cole, S. & Frenk, C. S. Cluster evolution as a diagnostic for omega. *Mon. Not. R. Astron. Soc.* **282**, 263–280 (1996).

Acknowledgements. We thank S. White for discussions. The simulations were performed at ARSC and NCSA supercomputing centers. This work was supported by the EU Network for Galaxy Formation and Evolution, the NASA/ESS programme, and PPARC. C.S.F. acknowledges a PPARC Senior Research Fellowship and S.C. a PPARC Advanced Fellowship; C.G.L. was supported by the Danish National Research Foundation through its establishment of The Theoretical Astrophysics Center.

Correspondence and requests for materials should be addressed to C.S.F. (e-mail: c.s.frenk@durham.ac.uk).

Negative Poisson's ratios as a common feature of cubic metals

Ray H. Baughman*, Justin M. Shacklette*, Anvar A. Zakhidov* & Sven Stafström†

* AlliedSignal Inc., Research and Technology, Morristown, New Jersey 07962-1021, USA

† Department of Physics and Measurement Technology, Linköping University, S-581 83, Linköping, Sweden

Poisson's ratio is, for specified directions, the ratio of a lateral contraction to the longitudinal extension during the stretching of a material. Although a negative Poisson's ratio (that is, a lateral extension in response to stretching) is not forbidden by thermodynamics, this property is generally believed to be rare in crystalline solids¹. In contrast to this belief, 69% of the cubic elemental metals have a negative Poisson's ratio when stretched along the [110] direction. For these metals, we find that correlations exist between the work function and the extremal values of Poisson's ratio for this stretch direction, which we explain using a simple electron-gas model. Moreover, these negative Poisson's ratios permit the existence, in the orthogonal lateral direction, of positive Poisson's ratios up to the stability limit of 2 for cubic crystals. Such metals having negative Poisson's ratios may find application as electrodes that amplify the response of piezoelectric sensors.

There is considerable fundamental and practical interest in materials with negative Poisson's ratios^{1–14}, which Evans⁴ called auxetic. We refer to auxetic crystals as axially auxetic or non-axially auxetic, depending upon whether or not a negative Poisson's ratio arises for a crystal-axis direction. The calculation of Poisson's ratio is complicated for directions oblique to the crystal axes, because the elastic constant tensor for general orientations involves as many as 21 interrelated components for even a cubic phase¹⁵. This complication probably explains why the common existence of non-axial auxetic behaviour has largely remained unrecognized. However, the important early work of Milstein and Huang¹⁶ established the occurrence of non-axial auxetic behaviour for cubic metals and rare gases. The origin of this behaviour must differ from that of other known types of auxetic materials and structures: foams, honeycombs and hypothetical carbon phases having re-entrant or hinged structures^{2–5,8,13}; microporous organic polymers⁴; polymer laminates^{11,14}; polymer/fibre composites¹²; and crystals such as α -cristobalite^{1,6}.

We simplify the search for auxetic materials by deriving a general criterion for the existence of auxetic behaviour. This 'auxetic criterion', expressed in elastic compliances (S_{ij}) for the axial directions, is that $S_{11} + S_{33} + 2S_{13} - S_{44} > 0$. Satisfying this inequality is a necessary and sufficient condition for auxetic behaviour for a hexagonal or cubic phase when the Poisson's ratios for axial directions are all positive. This general criterion is derived from equations for the Poisson's ratio for an arbitrary direction for cubic¹⁵ and hexagonal¹⁷ phases. Using these equations, the Poisson's ratios that usually have extremal values for cubic phases are:

$$\nu(110, 1\bar{1}0) = -(2C_{11}C_{44} - (C_{11} - C_{12})(C_{11} + 2C_{12}))/ (2C_{11}C_{44} + (C_{11} - C_{12})(C_{11} + 2C_{12})) \quad (1)$$

$$\nu(110, 001) = 4C_{12}C_{44}/(2C_{11}C_{44} + (C_{11} - C_{12})(C_{11} + 2C_{12})) \quad (2)$$

These are the Poisson's ratios for a [110] stretch, measured for $[1\bar{1}0]$ and $[001]$ lateral directions, respectively. Such Poisson's ratios are

for a stretch along the face diagonal of the cubic cell and a resulting lateral strain measured along a perpendicular face-diagonal and a perpendicular cube-axis direction, respectively.

We identified non-axial auxetic behaviour by the application of the above criterion to experimentally determined elastic compliance matrices (see ref. 18 and references therein, and Supplementary Information). About 69% of the 32 investigated cubic phases of the elemental metals satisfy the auxetic criterion. Of these 18 face-centred cubic (f.c.c.) metals and 14 body-centred cubic (b.c.c.) metals, 78% and 57%, respectively, are non-axially auxetic. For all of these cubic phases, except lithium, both the minimum and the maximum Poisson's ratio are for a [110] applied stress. In contrast, out of 20 investigated hexagonal close-packed (h.c.p.) phases, only

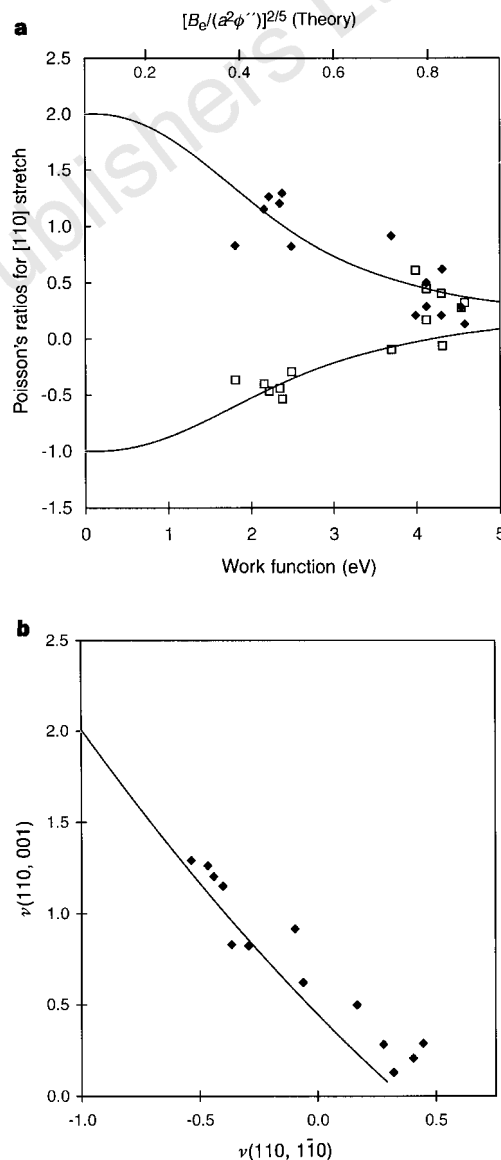


Figure 1 Correlations for the Poisson's ratios of b.c.c. metals, derived from observed^{18,22} elastic compliances and polycrystalline work functions. **a**, Correlation between $\nu(110, 1\bar{1}0)$ and $\nu(110, 001)$ and the polycrystalline work function. The experimentally derived Poisson's ratios are denoted by unfilled squares for $\nu(110, 1\bar{1}0)$ and filled diamonds for $\nu(110, 001)$. The curves are predicted dependencies of $\nu(110, 1\bar{1}0)$ and $\nu(110, 001)$ (bottom and top curves, respectively) on $[B_e/(a^2\phi'')]^{2/5}$; see text for details. **b**, Comparison of experimentally derived data points relating $\nu(110, 1\bar{1}0)$ and $\nu(110, 001)$ with the theoretical prediction (solid curve) for the simple two-term lattice energy used to obtain the theoretical curves of **a**.

zinc and beryllium are auxetic. Other than zinc, none of these cubic and hexagonal phases of elemental metals has a reliably reported negative Poisson's ratio for an axial direction. These results indicate for the elemental metals that (1) f.c.c. and b.c.c. phases are usually auxetic in non-axial directions, (2) axial auxetic behaviour is rare (or non-existent) for the cubic phases, and (3) the hexagonal phases are rarely auxetic for either axial or non-axial directions. None of the cubic phases are auxetic for a stretch along the three-fold axis direction, which is also the case for the auxetic trigonal phases of arsenic and bismuth characterized by Gunton and Saunders¹⁹.

Our analysis of reported elastic compliance data for cubic alloys and the cubic intermetallic phases¹⁸ provides results consistent with the above conclusions. Although none of the ~130 investigated cubic alloy and cubic intermetallic phases are auxetic for axial directions, ~50% of these phases are auxetic for non-axial directions. Furthermore, all those cubic alloy and cubic intermetallic phases that are auxetic contain at least one element that has an auxetic phase, and auxetic behaviour is almost always found for such cubic phases that contain only elements having auxetic phases.

In contrast with these results for metals, application of the auxetic criterion to published elastic compliance matrices¹⁸ shows that auxetic behaviour of any type is either rare or non-existent for many classes of cubic materials. For example, all investigated cubic alums (48 phases) and garnets (20 phases) are non-auxetic. Also, auxetic behaviour is rare or non-existent for salts with either the NaCl or the CsCl type structure. The diamond structural types for carbon, silicon and germanium are also non-auxetic. However, various cubic solids that share the ball-like packing of the elemental metals are auxetic. For example, non-axial auxetic behaviour is found^{16,18,20,21} for the f.c.c. phases of rare gases, methane, ammonia and adamantane (but not for the f.c.c. phase of C₆₀).

The existence of negative Poisson's ratios for the cubic metals enables the existence of Poisson's ratios above unity. In fact, our analysis of the literature data¹⁸ for metallic cubic phases provides non-axial Poisson's ratios that range between -0.67 and 1.47 for Cu_{68.6}Al_{27.6}Ni_{3.8} and between -0.81 and 1.68 for CuAuZn₂. Also, ~20% of the ~80 investigated types of cubic alloys provide a maximum Poisson's ratio of above unity. Such metals must have a negative Poisson's ratio for the same stretch direction as the

maximum Poisson's ratio ($\nu_{\max}$), otherwise the linear compressibility would be negative (thereby violating a stability condition for cubic phases). The existence of a $\nu_{\max} > 1$ implies that a plane exists parallel to the stretch axis that decreases area during stretching. The most negative fractional area change ($\Delta A/A$) caused by a fractional elongation of $\Delta L/L$ is then $\Delta A/A = (1 - \nu_{\max})\Delta L/L$.

Figures 1 and 2 show results that can be used both to understand the origin of the negative Poisson's ratios and to design new auxetic phases. These figures for b.c.c. metal phases (Fig. 1a) and for non-ferromagnetic f.c.c. metal phases (Fig. 2) show that the extremal Poisson's ratios for a [110] stretch are correlated with the measured work function²². We also find that correlations exist between either $\nu(110, \bar{1}\bar{1}0)$ or $\nu(110, 001)$ and $(n_{\text{ws}})^{1/3}$, $(H_{\text{vap}}/(V_{\text{M}})^{2/3})^{1/3}$ and $V_{\text{M}}^{1/3}$, where n_{ws} is the Wigner-Seitz electron density parameter²³, H_{vap} is the molar heat of vaporization, and V_{M} is the molar volume. These relationships reflect the correlation between off-axis Poisson's ratios and work function, as it is known²³ for metals that $(n_{\text{ws}})^{1/3}$, $(H_{\text{vap}}/(V_{\text{M}})^{2/3})^{1/3}$ and $V_{\text{M}}^{1/3}$ are correlated with work function.

We find that a simplistic model provides insight into the origin of the negative Poisson's ratios and the observed correlations. We approximate the energy per atom (U) of a crystal (with molar volume V and equilibrium molar volume V_0) by the sum of nearest-neighbour central force interactions (Φ) and a term that depends on the m th power of volume. The second derivative of this second term with respect to V/V_0 (evaluated at $V = V_0$) is the equilibrium value of the bulk modulus of the electron gas (B_e). Hence, $U = (p/2)\Phi + B_e(V/V_0)^m/(m(m-1))$, where p is the number of nearest-neighbour atoms surrounding a central atom. As the basic conclusions of the analysis do not depend on this approximation, we ignore core exclusion, exchange, and correlation effects, and approximate B_e by the bulk modulus due to the kinetic energy of the electron gas (so that m is $-2/3$)²⁴⁻²⁶. Following the approach used by Thomas²⁴, calculating the derivatives for elastic stiffnesses and applying the condition for equilibrium ($\delta U/\delta V = 0$ for $V = V_0$) results in $C_{11} - C_{12} = 2B_e(1-m)^{-1}$, $C_{44} = (a^2\Phi'' + 2B_e(1-m)^{-1})/3$, and $C_{11} + 2C_{12} = a^2\Phi'' + B_e(1-3m)(1-m)^{-1}$ for b.c.c. phases and $C_{11} - C_{12} = (a^2\Phi'' + 7B_e(1-m)^{-1})/4$, $C_{44} = (a^2\Phi'' + 3B_e(1-m)^{-1})/4$, and $C_{11} + 2C_{12} = a^2\Phi'' + B_e(1-3m)(1-m)^{-1}$ for f.c.c. phases, where Φ'' is $d^2\Phi/dr^2$ evaluated at the equilibrium interatomic separation $r = r_0$ and where a is the lattice parameter. These equations are substituted into equations (1) and (2) to obtain $\nu(110, \bar{1}\bar{1}0)$ and $\nu(110, 001)$ as a function of only $B_e/(a^2\Phi'')$. For nearly vanishing values of B_e , these equations provide minimum and maximum Poisson's ratios of -1 and 2 for b.c.c. and 0 and 1/2 for f.c.c. Simple geometrical arguments, which are illustrated in Fig. 3 for b.c.c., explain why nearest-neighbour central forces result in a negative Poisson ratio for b.c.c. phases, but not for f.c.c. or h.c.p. phases.

In Fig. 1b we compare the relationship between $\nu(110, \bar{1}\bar{1}0)$ and $\nu(110, 001)$ obtained from theory and experiment for b.c.c. phases. Considering the simplicity of this theory and the absence of any adjustable parameters, the agreement is quite good. Also, Fig. 1a shows that the theoretically predicted dependence of $\nu(110, \bar{1}\bar{1}0)$ and $\nu(110, 001)$ on $[B_e/(a^2\Phi'')]^{2/5}$ is quite similar to the experimentally determined variation of these Poisson's ratios with work function. These dependent variables were chosen for comparison of theory and experiment because free-electron theory predicts that $B_e^{2/5}$ is proportional to the Fermi energy (E_F), and the work function (W_F) differs in magnitude from E_F by only the contribution from surface dipoles. Hence increasing W_F increases the bulk modulus contribution from the electron gas. The above equations for b.c.c. metals imply that the auxetic behaviour disappears as the mechanical anisotropy becomes small, which is consistent with observations²¹. However, other types of structures are known that are both auxetic and isotropic¹⁴.

Comparison of the theoretical and observed dependence of the Poisson's ratios on work function for f.c.c. metals (Fig. 2) indicates

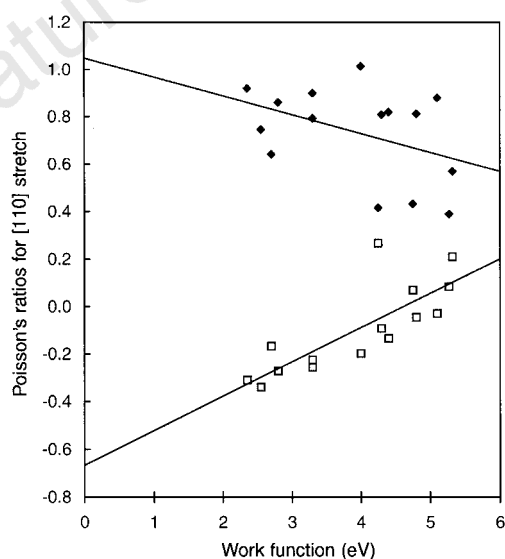


Figure 2 The dependence of $\nu(110, \bar{1}\bar{1}0)$ and $\nu(110, 001)$ on polycrystalline work function for f.c.c. phases of the non-ferromagnetic elemental metals. Unfilled squares denote $\nu(110, \bar{1}\bar{1}0)$ and filled diamonds denote $\nu(110, 001)$. The lines are least-squares fits to the experimental results, which are derived from reported compliances¹⁸ and work functions²².

qualitatively similar trends. But although the model predicts for b.c.c. that $\nu(110, 1\bar{1}0)$ approaches -1 and $\nu(110, 001)$ approaches 2 as B_e vanishes, the corresponding calculated Poisson's ratios for the f.c.c. phases are 0 and $1/2$, respectively. Hence, the present simplified theory must be modified for the f.c.c. metals by the inclusion of interatomic potential or band-structure effects that can be expressed as many-body forces. Evidence that such forces cannot be ignored for f.c.c. metals is provided by Cousins and Martin²⁷, who show that the crystal potential for f.c.c. metals like Cu, Ag, Au and Ni cannot be reliably expressed as the sum of pairwise interatomic and volume-dependent terms. Inclusion of many-body forces can shift the Poisson's ratios downwards to approximately the observed values. For example, application of the simple three-atom, Born–Mayer-type potential of Cousins²⁸ provides limiting values of $\nu(110, 1\bar{1}0) = -1$ and $\nu(110, 001) = 2$, which are identical to the results for b.c.c. phases that are described by the pairwise interaction model. As for the case of b.c.c. phases, such large-magnitude Poisson's ratios are reduced by the effect of the electron-gas bulk modulus, which increases with increasing work function.

Based on these results we make the following conclusions. Pairwise central forces can explain the large negative values of $\nu(110, 1\bar{1}0)$ and the large positive values of $\nu(110, 001)$ for the b.c.c. metals. An increasing electron-gas contribution to bulk modulus with increasing work function, which helps stabilize the

b.c.c. phases as an energy minima, simultaneously makes $\nu(110, 1\bar{1}0)$ less negative and $\nu(110, 001)$ less positive. The dependence of $\nu(110, 1\bar{1}0)$ and $\nu(110, 001)$ on work function is similar for the f.c.c. and b.c.c. metals. However, significant auxetic behaviour for the f.c.c. metals depends on the existence of a substantial many-body contribution to the binding energy. The present analysis indicates that Poisson's ratios of $\nu(110, 1\bar{1}0) = -1$ and $\nu(110, 001) = 2$ can be closely approached by either f.c.c. or b.c.c. phases that have negligible electron-gas contributions—although precisely reaching these values results in a soft mode shear instability. The above conclusions from theory and experimental results are opposite to the previous suggestion²¹ that the electron-gas contribution causes negative Poisson's ratios.

The existence of negative Poisson's ratios for the cubic metals has practical importance, especially as a negative Poisson's ratio reverses the compensation effects of positive Poisson's ratios on the volume and area changes caused by a uniaxial stress. One example is a method for obtaining over a two-fold increase in the sensitivity of a strain sensor by sandwiching a sheet of piezoelectric polymer between two thick auxetic metal electrodes. This sensitivity increase occurs because the negative Poisson's ratio of the metal electrode amplifies the effect of an applied in-plane uniaxial strain on sensor sheet area.

Some types of applications previously proposed for axially auxetic materials^{3,4} are also possible for the non-axially auxetic cubic metals. The prospects for such applications are enhanced by the commercial availability of non-axially auxetic metals as large single crystals. For example, large single crystals of Ni_3Al , which has¹⁸ a minimum Poisson's ratio of -0.18 , are used as vanes for aircraft gas turbine engines²⁹. $\square$

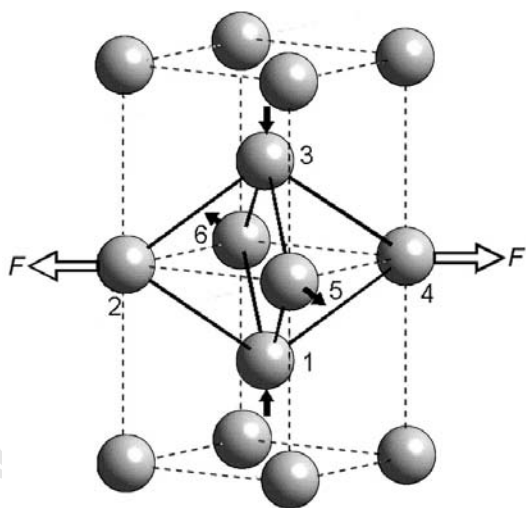


Figure 3 The structural origin of a negative Poisson's ratio and a giant positive Poisson's ratio for the case of a rigid-sphere b.c.c. solid. The solid minimizes density decrease during deformation by maintaining intersphere contact. Two b.c.c. unit cells (defined by the dotted lines) provide a crystallographic reference for describing the relative directions of atomic displacements (black arrows) of atoms 1–6 in response to an applied force in the $[110]$ direction (F , indicated by the two white arrows). Decreasing the acute angle in the rhombus defined by atoms 1–4 is the only way to elongate the crystal in the $[110]$ stress direction without increasing the nearest-neighbour separations (indicated by the solid black lines). Hence, the separation between atoms 1 and 3 decreases, providing a positive $\nu(110, 001)$. This angle decrease partially closes this rhombus, thereby pushing atoms 5 and 6 apart (corresponding to a negative $\nu(110, 1\bar{1}0)$). Simple geometrical analysis of these deformations provides that $\nu(110, 001) = 2$ and $\nu(110, 1\bar{1}0) = -1$. Similarly, simple pictures indicate that the minimum Poisson's ratio is zero for rigid-sphere f.c.c. and h.c.p. phases. Atoms in next-nearest-neighbour ($1\bar{1}0$) sheets are orthogonally in contact for f.c.c. Hence, the dimension in the $[1\bar{1}0]$ direction is unaffected by a $[110]$ strain and $\nu(110, 1\bar{1}0) = 0$. Likewise, an infinitesimal uniaxial elongation orthogonal to a close-packed plane of rigid-sphere atoms in h.c.p. cannot significantly affect packing in this plane, so the minimum Poisson's ratio is zero.

Received 7 April; accepted 1 December 1997.

- Keskar, N. R. & Chelikowsky, J. R. Negative Poisson ratios in crystalline SiO_2 from first-principles calculations. *Nature* **358**, 222–224 (1992).
- Gibson, L. J., Ashby, M. F., Schajer, G. S. & Robertson, C. I. The mechanics of two-dimensional cellular materials. *Proc. R. Soc. Lond. A* **382**, 25–42 (1982).
- Lakes, R. S. Foam structures with a negative Poisson's ratio. *Science* **235**, 1038–1040 (1987).
- Evans, K. E., Nkansah, M. A., Hutchinson, I. J. & Rogers, S. C. Molecular network design. *Nature* **353**, 124 (1991).
- Rothenburg, L., Berlin, A. A. & Bathurst, R. J. Microstructure of isotropic materials with negative Poisson's ratio. *Nature* **354**, 470–472 (1991).
- Yeganeh-Haeri, A., Weidner, D. J. & Parise, J. B. Elasticity of α -cristobalite—a silicon dioxide with negative Poisson's ratio. *Science* **257**, 650–652 (1992).
- Wei, G. & Edwards, S. F. Polymer networks with negative Poisson's ratios. *Comput. Poly. Sci.* **2**, 44–54 (1992).
- Baughman, R. H. & Galvão, D. S. Crystalline networks with unusual predicted mechanical and thermal properties. *Nature* **365**, 735–737 (1993).
- Smith, W. A. in *Proc. 1991 Ultrasonics Symp.* (ed. McAvoy, B. R.) 661–666 (IEEE, Piscataway, NJ, 1992).
- Wojciechowski, K. W. Two-dimensional isotropic solid with a negative Poisson's ratio. *Phys. Lett. A* **137**, 60–64 (1989).
- Lakes, R. S. No contractile obligations. *Nature* **358**, 713–714 (1992).
- Miki, M. & Morotsu, Y. The peculiar behavior of the Poisson's ratio of laminated composites. *JSMSE Int. J.* **32**, 67–72 (1989).
- Lakes, R. S. Deformation mechanisms of negative Poisson's ratio materials: structural aspects. *J. Mater. Sci.* **26**, 2287–2292 (1989).
- Milton, G. W. Composite materials with Poisson's ratios close to -1 . *J. Mech. Phys. Solids* **40**, 1105–1137 (1992).
- Turley, J. & Sines, J. The anisotropy of Young's modulus, shear modulus and Poisson's ratio in cubic materials. *J. Phys. D* **4**, 264–271 (1971).
- Milstein, F. & Huang, K. Existence of a negative Poisson ratio in fcc crystals. *Phys. Rev.* **19**, 2030–2033 (1979).
- Li, Y. Anisotropic behavior of Poisson's ratio, Young's modulus, and shear modulus in hexagonal materials. *Phys. Status Solidi A* **38**, 171–175 (1976).
- Every, A. G. & McCurdy, A. K. in *Landolt-Börnstein* (ed. Nelson, D. E.) 11–133 (New Ser. III/29a, Springer, Berlin, 1992).
- Gunton, D. J. & Saunders, G. A. The Young's modulus and Poisson's ratio of arsenic, antimony, and bismuth. *J. Mater. Sci.* **7**, 1061–1068 (1972).
- Fioretto, D. et al. Brillouin-scattering determination of the elastic constants of epitaxial fcc C_{60} film. *Phys. Rev. B* **52**, R8707–R8710 (1995).
- Jain, M. & Verma, M. P. Poisson's ratios in cubic crystals corresponding to (110) loading. *Ind. J. Pure Appl. Phys.* **28**, 178–182 (1990).
- Grigoriev, I. S. & Melnikov, E. Z. (eds) *Handbook of Physical Properties* Ch. 25, 697–698 (CRC Press, New York, 1997).
- Miedema, A. R., de Châtel, P. F. & de Boer, F. R. Cohesion in alloys—fundamentals of a semi-empirical model. *Physica B* **100**, 1–28 (1980).
- Thomas, J. M. Failure of the Cauchy relation in cubic metals. *Scripta Metall.* **5**, 787–790 (1971).
- Gilman, J. J. Bulk stiffnesses of metals. *Mater. Sci. Eng.* **7**, 357–361 (1971).
- Ashcroft, N. A. & Mermin, N. D. *Solid State Physics* Ch. 2, 30–55 (Holt, Rinehardt & Winston, New York, 1976).
- Cousins, C. S. G. & Martin, J. W. Extended Cauchy discrepancies—measures of non-central, non-isotropic interactions in crystals. *J. Phys. F* **8**, 2279–2291 (1978).

28. Cousins, C. S. G. Evidence for short range, three-body interactions in copper. *J. Phys. F* **3**, 1915–1920 (1973).
 29. Nakamura, M. Fundamental properties of intermetallic compounds. *Mater. Res. Soc. Bull.* **XX**(8), 33–39 (1995).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. We thank R. S. Lakes, J. J. Gilman, L. W. Shacklette, R. C. Morris, and S. O. Dantas for important comments.

Correspondence and requests for materials should be addressed to R.H.B. (e-mail: ray.baughman@alliedsignal.com).

Voltage-dependent anchoring of a nematic liquid crystal on a grating surface

G. P. Bryan-Brown, C. V. Brown, I. C. Sage & V. C. Hui

Defence Evaluation and Research Agency, St Andrews Road, Malvern, Worcestershire WR14 3PS, UK

The switching properties of most liquid-crystal electro-optic devices rely mainly on the reorientation of the average molecular direction (director) within the bulk of the liquid-crystal layer¹. Reorientation of the director at or near the surfaces of the layer usually has an insignificant effect on device performance. Here we describe a different configuration in which a nematic liquid crystal is placed between a flat surface treated to induce a parallel anchoring of the director and a grating surface treated to give a perpendicular anchoring. We show that this configuration leads to an effective azimuthal anchoring at the grating surface that depends on the applied voltage when the nematic phase has negative dielectric anisotropy (that is, the director has a tendency to align perpendicular to the applied field). This leads to a voltage-controlled twist effect in the liquid-crystal cell that is highly sensitive to the grating profile. Furthermore, this twist effect possesses an electro-optic response which is far less dependent on viewing angle compared to many other liquid-crystal display configurations. We therefore suggest that this technology might find application in the next generation of liquid-crystal displays.

Liquid-crystal displays consist of a liquid-crystal layer contained between two substrate surfaces. These surfaces impose a preferred orientation of the liquid-crystal director to generate uniform optical properties. Treating the surface with a low-surface-energy surfactant² leads to perpendicular alignment (director orthogonal to surface plane) whereas an ordered polymer coating leads to planar alignment (director in surface plane with preferred direction). Planar alignment is usually achieved using a polymer which has been ordered by rubbing with a nylon cloth³, although prototypes are now emerging in which the ordering is induced optically⁴. The earliest alignment method to be explained quantitatively is that in which a grating surface is used⁵. The liquid-crystal director aligns along the grating grooves in order to minimize the elastic distortion energy.

Here we consider a nematic liquid crystal in contact with a grating surface which has been treated with a low-energy surfactant. Such a treatment constrains the director to align perpendicular to the local surface direction. The average director orientation is then perpendicular to the average surface direction for a symmetric grating profile. If the liquid crystal has negative dielectric anisotropy, then application of a sufficient electric field will lead to a planar configuration in which the director lies along the grating grooves. This is also well known. However, in the present work the grating surface is placed opposite a normally planar surface in which the planar anchoring direction is perpendicular to the grating groove

direction. A new mode of operation is then observed. Figure 1 shows a schematic view of the behaviour of this configuration in which the grating grooves are in the y direction (that is, out of the plane of the figure) and the planar anchored direction is along the x axis. The director is represented by the black rods. When no voltage is applied between the substrates ($V = 0$), the director has a tilt in the x – z plane which varies roughly linearly with z . When the voltage is raised to V_1 , the tilt angle decreases because the applied field couples to the negative dielectric anisotropy but the director remains in the x – z plane as this is the preferred anchoring direction on the planar surface. However, when the voltage is increased the liquid-crystal tilt is lowered in the vicinity of the grating surface which then starts to impose an azimuthal twisting torque on the near-surface director. Eventually, at a sufficient voltage (V_2), the twist torque overcomes the bulk twist threshold energy and a finite twist is observed. These energies may be compared as follows. The maximum twist torque which can be imposed by the grating is given by⁵

$$W_\phi = \frac{1}{4} \sqrt{k_{11}k_{33}} a^2 \left(\frac{2\pi}{L} \right)^3 \quad (1)$$

where a and L are respectively the amplitude and pitch of a sinusoidal oscillation, and k_{11} and k_{33} are the liquid-crystal elastic constants. For a typical grating with an amplitude of $0.25 \mu\text{m}$ and a pitch of $1 \mu\text{m}$, this energy is roughly $4 \times 10^{-5} \text{ J m}^{-2}$. The twist threshold energy is given by k_{22}/d where d is the cell gap. For a $5\text{-}\mu\text{m}$ cell this energy is typically only $1 \times 10^{-6} \text{ J m}^{-2}$. Therefore in the high-voltage regime, the grating alignment energy dominates over the twist energy and the cell is expected to have a 90° twist as shown in Fig. 1 at $V = V_2$. This configuration thus allows a voltage-controlled twist (VCT) effect.

The VCT behaviour can also be numerically analysed using a finite-element approach in which a grid of directors are elastically coupled to each other and contained between surfaces with defined boundary conditions. For a particular applied voltage, the system is relaxed until the minimum energy is found. Figure 2a shows an example of a predicted configuration at 0 V . In this case, the liquid crystal is continuously distorted between the planar boundary condition on the top flat surface and the perpendicular boundary condition on the bottom grating surface. If the voltage is set to 15 V (Fig. 2b), the cell relaxes into a configuration with lower average tilt in which the director near the grating is twisted out of the x – z plane (represented by shorter lines). Therefore this approach also predicts a VCT effect.

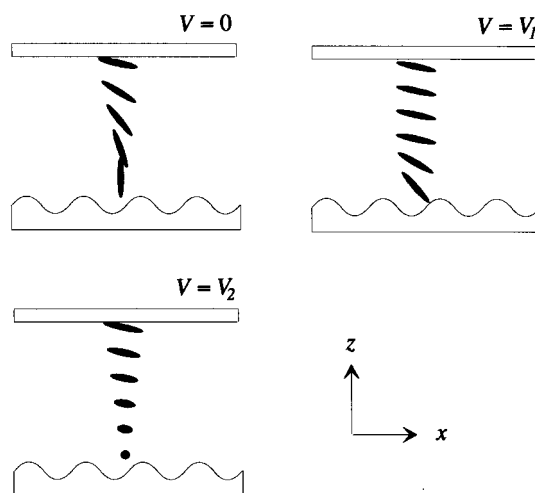


Figure 1 Schematic representation of the liquid-crystal cell geometry used to observe the VCT effect. Details are given in the text.

Experimental samples were made in which the grating surface was defined in a photoresist layer (Shipley 1805, Shipley Europe Ltd, Coventry, UK) using hard contact photolithography. The groove pitch was defined by the spacing of the metal lines on the mask while the groove depth could be easily adjusted by changing the photo-exposure time. After processing, the surface was coated with a chrome complex coupling surfactant² in order to achieve the perpendicular boundary condition. Liquid-crystal cells with a 5- μm gap were then made using one grating surface and one planar (rubbed polymer) surface. Both substrates had an indium tin oxide underlayer in order to allow an electric field to be applied across the liquid-crystal layer. Cells were filled with the nematic liquid crystal MLC 6608 (from Merck) in its isotropic phase (at 92 °C) and then slowly cooled to room temperature.

Figure 3a shows data taken from a typical VCT cell placed between crossed polarizers along the x and y directions. In this case the grating pitch is 1 μm and the groove depth is 0.5 μm . The solid line shows the normalized white-light transmission as a function of applied voltage (1 kHz a.c.). The transmission is small (<0.5%) until 2.3 V when a steep rise is obtained which eventually saturates at a value of 92%. This quantity depends on the total amount of twist in the liquid-crystal layer. The dotted line shows the capacitance across the liquid-crystal layer which rises from 0 V without a threshold. The capacitance is sensitive to the average tilt angle in the cell as the liquid crystal possesses dielectric anisotropy. The two curves suggest that the twist change is the dominant process in the electro-optic effect, as the rise in transmission is much steeper than the rise in capacitance.

The dependence of the twist threshold on the grating profile was demonstrated by comparing data from cells in which the groove

depth was varied. Figure 3b shows data taken from three different VCT cells: the inset shows the grating profile in each cell measured by atomic force microscopy. It can be seen that the optical threshold voltage is smallest for the grating with the deepest grooves, whereas shallow grooves lead to a high voltage threshold. All these cells were found to have a capacitance response similar to that shown in Fig. 3a. These results show that gratings with lower anchoring energies (given by equation (1)) require a lower tilt to be achieved in the vicinity of the grating surface before the effective grating anchoring energy is able to overcome the twist threshold energy. Different grating profiles can also be combined in the same cell to give an optical response which varies with position. Figure 4 shows the transmission through such a cell at four different applied voltages. Three different grating groove depths have been created using a modified photolithographic mask. Raising the voltage leads to each region being 'switched on' in turn.

The VCT mode is a rare example of a liquid-crystal optical effect which depends strongly on surface properties. The results in Fig. 4 show that a spatially varying optical response can be derived from a cell with a single connection electrode on each substrate, and so the device is operating as an electro-optic analogue to digital converter which could have many applications. The second, perhaps more important, aspect of the VCT effect is that the grating surface has converted an externally induced tilt torque in the x - z plane into a twist torque in the x - y plane. Thus the grating is acting as a 'torque converter'. Achieving an electro-optic response from a twist effect rather than from a tilt effect in a liquid-crystal device has been shown to lead to superior viewing characteristics but usually requires a complex arrangement of in-plane electrodes to allow field application in the x - y plane⁶. These in-plane switch devices are

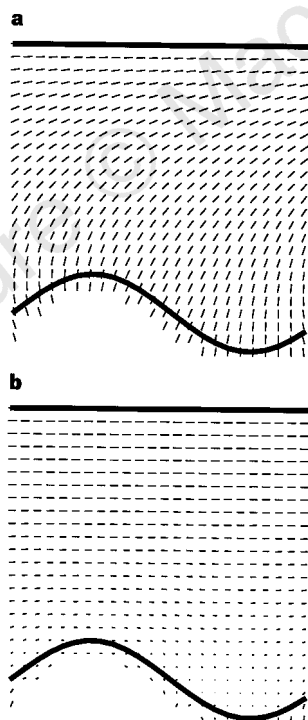


Figure 2 Finite-element modelling of the VCT configuration. The applied voltage is 0 V in **a**, and 15 V in **b**: the nematic liquid crystal has $k_{11} = 16.7$ pN, $k_{22} = 8.0$ pN, $k_{33} = 18.1$ pN and $\Delta\epsilon = -4.2$ where k_{11} , k_{22} and k_{33} are the splay, twist and bend elastic constants and $\Delta\epsilon$ is the dielectric anisotropy.

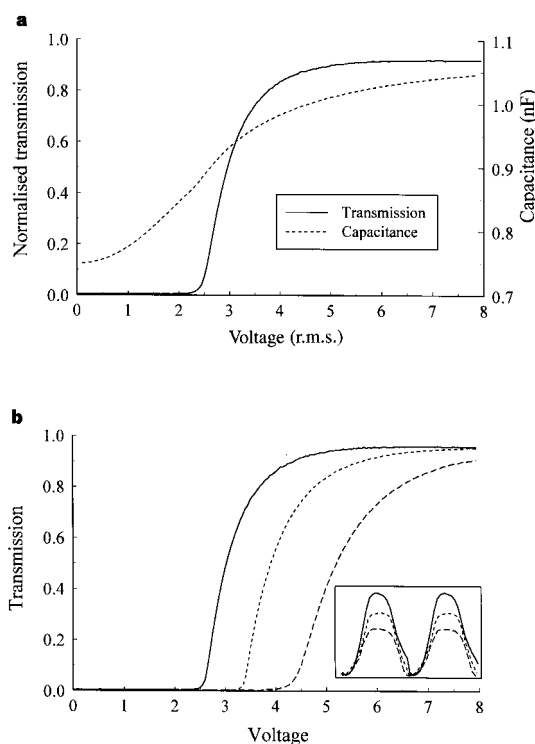


Figure 3 Voltage dependence of properties of VCT liquid-crystal cells. **a**, Results from a typical VCT liquid-crystal cell, showing transmission between crossed polarizers (solid line) and capacitance (dotted line) as a function of applied r.m.s. voltage (1 kHz). **b**, Transmission versus voltage for three cells containing gratings of different groove depths. Inset shows the actual grating profiles measured by atomic force microscopy. The deepest groove depth is 0.514 μm and the pitch of all the samples is 1.0 μm .

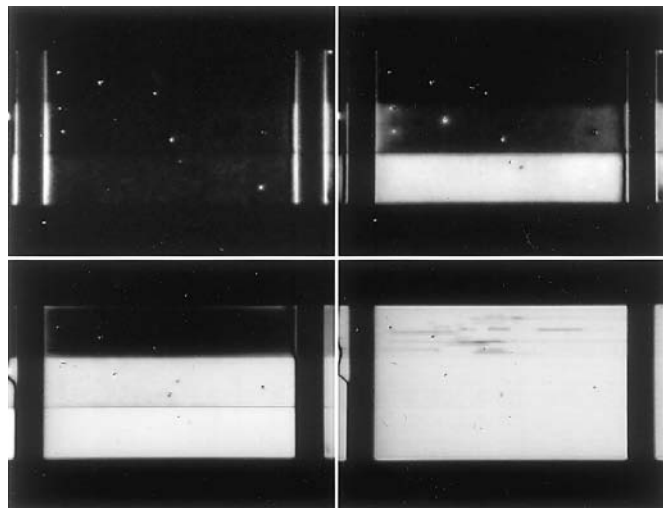


Figure 4 Photographs of a VCT cell showing voltage-dependent transmission. Applied voltages are 2.08 V (top left), 2.57 V (top right), 3.64 V (bottom left) and 8.14 V (bottom right).

at present attracting much interest in the displays industry. Using the VCT effect we can mimic in-plane switching without using in-plane electrodes, which greatly simplifies manufacture. So far results have shown that the VCT maintains an excellent contrast ratio for all viewing angles in the x - z plane. Viewing in the x - y plane is compromised by the partially tilted director profile at low voltage, but this can be overcome by adjusting the grating profile to induce an electro-optic response at higher voltages. Optical diffraction at the grating interface does lead to a slight loss of throughput (<5%) but this can be reduced by decreasing the grating pitch and by optimizing the refractive index of the grating material. The VCT mode only suffers a small change in capacitance across the operating voltage range (see Fig. 3a) compared to the in-plane switch mode which is likely to lead to a reduction in electrical power consumption. A typical VCT optical response time ($\tau_{\text{on}} + \tau_{\text{off}}$ (10%–90%), 0–5 V, MLC6608 in a 4.7- μm cell) is 80 ms which is much faster than the in-plane switch mode operating at the same voltage⁶.

We note that the in-plane switch mode can only be used in displays which incorporate thin-film transistors at each pixel, whereas the VCT mode can also be used in 'passive matrix' displays which consist of simple row and column electrodes. At present the VCT mode has an electro-optic steepness similar to a conventional twisted nematic display but optimization of elastic-constant ratios could lead to a much steeper response which would allow highly multiplexed passive matrix displays. If the problem of large area grating patterning can be solved then the VCT is likely to find many applications in future display devices, especially if further optimization can be achieved. □

Received 10 November 1997; accepted 9 January 1998.

- Shanks, I. A. The physics and display applications of liquid crystals. *Contemp. Phys.* **23**, 65–91 (1982).
- Cognard, J. Alignment of nematic liquid crystals in their mixtures. *Mol. Cryst. Liq. Cryst.* (Suppl. Ser.) **A 5**, 1–77 (1982).
- Geary, J. M., Goodby, J. W., Kmetz, A. R. & Patel, J. S. The mechanism of polymer alignment of liquid crystal materials. *J. Appl. Phys.* **62**, 4100–4108 (1987).
- Schadt, M., Seiberle, H. & Schuster, A. Optical patterning of multidomain liquid-crystal displays with wide viewing angles. *Nature* **381**, 212–215 (1996).
- Berberman, D. W. Alignment of liquid crystals by grooved surfaces. *Mol. Cryst. Liq. Cryst.* **23**, 215–231 (1973).
- Masahito, O. & Kondo, K. The in-plane switching of homogeneously aligned nematic liquid crystals. *Liq. Cryst.* **22**, 379–390 (1997).

Acknowledgements. We thank A. Smout for AFM measurements.

Correspondence and requests for materials should be addressed to G.B.-B. (e-mail: gpbryanbrown@DERA.GOV.UK).

Polymeric emulsifiers based on reversible formation of hydrophobic units

Arvind M. Mathur*, Bernhard Drescher*, Alec B. Scranton* & John Klier†

* Department of Chemical Engineering, Michigan State University, East Lansing, Michigan 48824, USA

† Central Research and Development, Dow Chemical Company, Midland, Michigan 48674, USA

Emulsions consist of mixtures of immiscible liquids where one liquid is finely dispersed within the continuous phase of another. They are generally not thermodynamically stable: the dispersion tends to separate over time. Aqueous emulsions, widely used in food, pharmaceutical, and many other industries, are often stabilized by block copolymers containing alternating hydrophilic and hydrophobic segments (typically based on ethylene oxide/propylene oxide diblock and triblock systems) that penetrate into the oil and aqueous phase, respectively^{1,2}. Here we describe a conceptually new type of emulsifier whose hydrophobic blocks are formed spontaneously and reversibly by the complexation of hydrophilic segments, thereby allowing the stabilizing properties of the system to be switched on and off. We illustrate this approach using a comb-type graft copolymer containing a poly(methacrylic acid) backbone and short grafts of poly(ethylene glycol). The uncomplexed polymer is hydrophilic, but acidic conditions induce the formation of hydrogen-bonded hydrophobic complexes between parts of the backbone and the grafts. As a result, the grafted copolymer forms alternating blocks of hydrophilic (uncomplexed) and hydrophobic (complexed) segments that stabilize acidic emulsions. An increase in pH suppresses complex formation and thus leads to the breakup of the emulsion. Emulsion tests show that although the performance of the grafted copolymers is not yet competitive with existing emulsifiers, this approach provides an efficient strategy for the design of fully reversible emulsifiers.

Complexes of poly(methacrylic acid) (PMAA) with free (ungrafted) poly(ethylene glycol) (PEG) have been well characterized in aqueous solution using a wide variety of physical, spectroscopic and compositional techniques (see refs 3–7 for reviews). It has been concluded that these polymers form hydrogen-bonded complexes which are stabilized by hydrophobic interactions in water. As polymer complexes are generally cooperative in nature, the stability of the complex increases with increasing molecular mass. Experiments performed with PMAA of high molecular mass revealed a critical PEG relative molecular mass (2,000) below which complexation does not occur^{3–7}. In addition, investigators have shown that complexes form in a 1:1 methacrylic acid (MAA) to ethylene glycol (EG) repeat unit ratio, and that the complexes can be broken by neutralization of the PMAA, or by the addition of alcohols. Owing to the hydrophobic stabilization, the strength of the complex increases with increasing temperature.

In previous work, we have characterized the complexation equilibrium in copolymers containing PMAA backbone chains and complementary PEG grafts^{8,9}. The grafted systems, like the ungrafted systems discussed above, have complexes with a 1:1 repeat unit ratio, and the complexes are disrupted by the addition of bases or alcohols. However, the grafted systems show some important differences as well. Covalently linking together the complementary PEG and PMAA constituents removes the critical chain length effect (complexes were formed for all relative molecular masses) and even highly diluted polymer complexes can be repeatedly broken and

reformed by cycling the system between complex-promoting and complex-breaking conditions⁸. In addition, the graft copolymers allow non-stoichiometric complexes to be readily prepared simply by using non-stoichiometric repeat-unit ratios during synthesis. In contrast, the free (ungrafted) systems tend to rearrange and form only stoichiometric complexes. A theoretical study of the complexation equilibrium revealed an entropic driving force for loop formation, especially for long grafts, and that several small loops are favoured over a single large loop⁹. The study also revealed that the conformational averages for the ungrafted (free) case asymptotically approach those for the grafted case as the segmental binding free energy, polymer concentration or graft length are increased.

We synthesized graft copolymers with a poly(methacrylic acid) backbone and oligo(ethylene glycol) grafts, as shown in Fig. 1. Under acidic conditions the PEG grafts may form 1:1 hydrogen-bonded complexes with the PMAA (shown pictorially in Fig. 2), and the complex is considerably more hydrophobic than the constituent PMAA and PEG homopolymers^{8,10,11}. Therefore, in the complexed state, the polymer resembles a hydrophobic/hydrophilic multi-block copolymer (see Fig. 1). As the polymer complexes can be reversibly disrupted and reformed by a change in pH, solvent type or temperature, the polymers may be reversibly switched between a completely hydrophilic graft copolymer state to a hydrophobic-hydrophilic multi-block copolymer state.

Two classes of copolymers were designed to have dramatically different emulsification properties in response to changes in pH. The first class of polymers was designed to produce stable emulsions under acidic conditions, but not under basic conditions. Our strategy for these copolymers was to use non-stoichiometric ratios of the complexing repeat units to ensure an amphipathic character under complex-promoting (acidic) conditions. Specifically, representative copolymers were synthesized with MAA to EG repeat unit ratios of 10:1 (copolymer **1a**) and 20:1 (copolymer **1b**). Under complex-promoting conditions, this molecular architecture leads to relatively short hydrophobic segments (~20 repeating

units), and much longer hydrophilic segments which may extend (or loop) into the aqueous phase to provide steric stabilization.

The second class of copolymers was designed to produce stable emulsions under basic conditions, but not under acidic conditions. In this case, our strategy was to use stoichiometric ratios of the MAA and EG repeat units to render the polymer completely hydrophobic (water-insoluble) under acidic (complex-promoting) conditions, and to include long-chain alkyl grafts so that the copolymer assumes an amphipathic nature under basic conditions (when the complex is disrupted). Therefore, under basic conditions the pendent hydrophobic segments are expected to partition onto the oil-water interface while the hydrophilic PMAA backbone and PEG grafts provide electrostatic and steric stabilization. Accordingly, copolymer **2** was synthesized using a 1:1 MAA to EG repeat unit ratio, with the addition of lauryl methacrylate (10 mol% of the MAA repeat units) to impart the hydrophobic grafts to the final polymer.

Figure 3 shows the pH-dependent emulsification properties of the copolymers **1a**, **1b** and **2**. This figure shows the percentage of oil emulsified as a function of pH for each of the copolymers, and for a PMAA homopolymer. It may be seen that, as designed, the PMAA-g-PEG copolymers (**1a** and **1b**) form stable emulsions under acidic conditions (100% of the oil is emulsified), but do not form stable emulsions under basic conditions. In contrast, the PMAA-g-PEG-co-LM copolymer (**2**) has an opposite trend, with 100% of the oil emulsified under basic conditions, and 0% emulsified under acidic conditions. The relatively sharp transition observed for copolymer **2** may arise from an enhanced cooperativity of the hydrophobic association for the entirely complexed (hydrophobic) chains. In addition, comparison of the emulsification curves of copolymers **1a**, **1b** and **2** with that of the PMMA homopolymer reveals the pivotal role of the reversible hydrophobes: only the block/graft copolymers lead to the formation of stable emulsions that can be rapidly broken on demand by the prescribed change in pH. The emulsification properties of the polymers are reversible, with the same emulsification curves being obtained on repeated cycling between acidic and basic conditions.

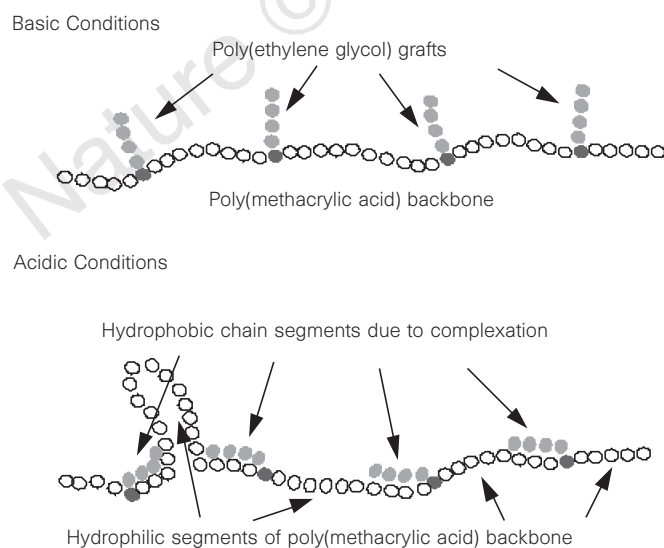


Figure 1 Schematic depiction of copolymer **1** under acidic (complex-promoting) conditions and under basic conditions. Although only loop-free, intramolecular complexes are shown here, both loops and intermolecular complexes are possible. Based on statistical considerations, loop formation is expected to be improbable for the relatively short grafts used in this study⁹. Similarly, due to the relatively short graft lengths and the dilute conditions, essentially all of the complexes formed initially will be intramolecular. After the emulsion is formed and the emulsifier is concentrated at the oil-water interface, some complexation exchange may occur, and intermolecular complexes may be formed.

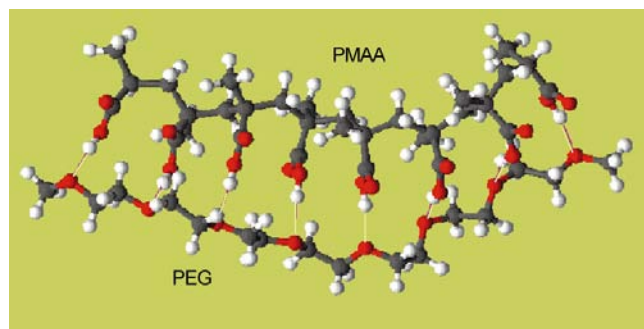


Figure 2 Representation of a 1:1 MAA:EG repeat unit ratio complex formed between PMAA and PEG segments, each containing eight repeat units. Hydrogen, carbon and oxygen atoms are shown as white, black and red, respectively. The hydrogen bonds are depicted by white lines between the ether oxygens and acid protons. This figure shows that it is sterically possible to form stable 1:1 complexes with minimal bond strain. This figure was prepared by performing molecular-mechanics simulations using the commercial software package POLYGRAF based upon the DRIEDING force field¹³. To achieve a realistic initial condition, the individual structures were energy-minimized and charge equilibrated before the molecular-mechanics studies were begun. The complementary polymer segments were placed in proximity, and energy-minimization calculations (molecular mechanics) were performed until an energy minimum was reached and the total energy calculations converged. The cut-off distance for hydrogen bonds between the ether oxygens and the acid protons was set at 3.5 Å.

The data in Fig. 3 also illustrate that within the first class of copolymers (**1a** and **1b**), the molecular architecture strongly influences the performance of the emulsifiers. For example, comparison of the curves for polymers **1a** and **1b** reveals that the percentage of oil emulsified depends upon the MAA to EG repeat unit ratio: although both polymers show the same general trend (oil is only emulsified under acidic conditions), the 20:1 MAA to EG copolymer (**1b**) shows a sharp decrease in emulsification capacity at a noticeably lower pH than the 10:1 (**1a**) system. This general trend probably arises from the greater number of hydrophilic (uncomplexed) MAA repeat units in the 20:1 copolymer (**1b**). Therefore, a lower degree of dissociation is required to remove the copolymer from the oil–water interface as the MAA to EG repeat units ratio is increased. This observation suggests that simple variations in molecular architecture may be used to tailor the pH range over which the emulsification capacities of the polymers change.

A detailed characterization of the emulsification mechanism will be published elsewhere. Briefly, the emulsion stabilization mechanism was attributed to stabilization against coalescence rather than facilitation of emulsification via interfacial tension reduction. This conclusion is based upon the following experimental observations: first, drop-volume interfacial tensiometry indicated that the emulsifiers did not reduce interfacial tension appreciably; second, material balances revealed that under emulsion-promoting conditions the polymers reside in the emulsion phase in the oil–water interfacial region. Third, an aqueous phase thickening mechanism was ruled out because rheological measurements revealed that maximum emulsion stability correlated with minimum aqueous-phase viscosity (viscosity changes were minimal at the concentrations of interest), and the oil droplets creamed rapidly through the low-viscosity (water-like) aqueous phase.

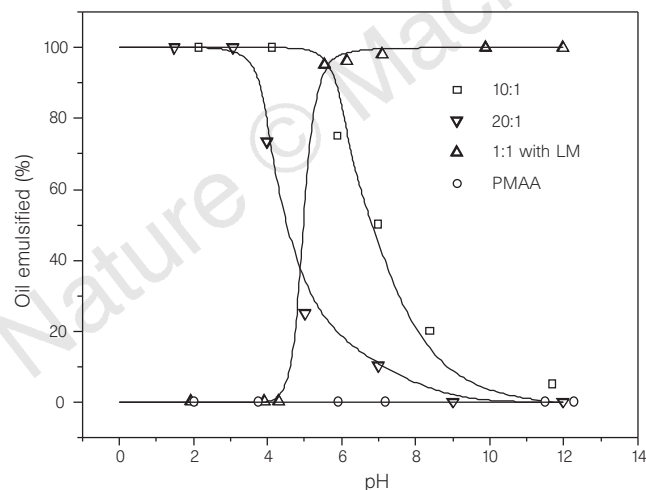


Figure 3 Percentage of oil emulsified after four days at room temperature as a function of pH. Results are shown for copolymers **1a** (10:1), **1b** (20:1), **2** (1:1 with lauryl methacrylate, LM), and MAA homopolymer. The emulsification properties of the copolymers were characterized using a standard bottle test method with methyl laurate as the model oil phase. This model oil is used widely as a solvent in biochemical and agricultural applications and serves as a replacement for mineral oil in many emulsifiable agricultural pesticide concentrates. All studies were carried out using 20 vol.% of the oil phase, and 80 vol.% of an aqueous phase containing 0.1 wt% of the surfactant. The pH of the aqueous phase was adjusted to the desired value by the addition of 1M HCl or 1M NaOH solutions. The mixture was agitated vigorously for ~2 min, then allowed to stand for four days at room temperature. The emulsified droplets formed an opaque, stable cream on top of the excess aqueous phase. The cream had a composition of ~70 vol.% oil in water, and individual droplets were not discernible by the naked eye. The percentage of oil emulsified was determined by a material balance after measurement of the volume of unemulsified oil.

The stability of emulsions prepared using a high-pressure homogenizer was characterized using laser confocal microscopy to measure the emulsion droplet-size distribution as a function of time. Emulsions containing 50 vol.% methyl laurate in water were prepared using a laboratory-scale high-pressure homogenizer (Biospec Products, Model No. 985–370) and 0.4 wt% of polymer **1a** as the emulsifier. The emulsions were maintained at pH 3.4, and the droplet size distribution was measured periodically (without diluting or perturbing the system) using a Zeiss 210 laser scanning microscope (Carl Zeiss, Thornwood, New York, USA). The methyl laurate phase was doped with trace quantities of the hydrophobic dye 3,3'-diiodatetradecyloxycarbocyanine (DiO). This fluorescent dye allowed oil droplets to be clearly distinguished from air bubbles in the aqueous phase (in fact very few air bubbles were observed). Images were collected in confocal fluorescence mode with excitation by the 488-nm line of an argon-ion laser, with the DiO fluorescence being detected using a 520–560 nm band-pass filter. A series of planar optical images (optical sections) were collected in both the x – y (parallel to the surface) and the x – z (perpendicular to the surface) planes, providing two orthogonal views of each droplet. Successive optical sections were separated by 1 μ m, and because the same droplets were followed in successive planar images, the true diameter of the droplets could be readily measured. Droplets with diameters smaller than 2.3 μ m could not be reliably measured and were disregarded; at least 50 distinct droplets were measured to characterize each distribution. The same sample was analysed in this manner one, four and ten days after preparation to characterize the evolution of the droplet size distribution. The experimental results showed that the droplet size distribution did not change significantly over time. After one, four and ten days, the number average size of the measured droplets was 6.14, 6.03 and 6.30 μ m, respectively. More significantly, the shape of the distribution did not change with time, and there was no generation of larger droplets over the course of ten days. These microscopic studies corroborate the macroscopic observation that the emulsions remain stable indefinitely.

However, the measured average droplet size of ~6 μ m indicates that existing emulsifiers are much better in producing fine droplets. (Conventional emulsifiers can produce microemulsions with droplets as small as 0.1 μ m.) We therefore expect that the new emulsifiers will find use in applications where the ability to form and break emulsions on demand is of particular value. Moreover, theoretical work suggests that multi-block copolymers may be more efficient as compatibilizers than di- or tri-block copolymers as the latter will lose more material into the bulk phase as micelles or mesophases¹². This feature suggests that further developments may yield emulsifiers with improved performance, particularly as the general approach described here could be extended to a wide range of other complexing systems that result in amphiphilic structures. □

Methods

The graft copolymers were synthesized by free-radical solution polymerization of methacrylic acid (MAA) with small amounts of an oligo(ethylene glycol) monomethacrylate macromonomer (MPEGMA-1000, containing an average of 22 ethylene glycol (EG) repeat units), and optionally, lauryl methacrylate. All polymerizations were carried out at 80 °C in a 1:1 (w/w) ethanol–water mixture (to ensure that complexes do not form during synthesis, and maintain solubility of all monomers) and were initiated using 0.1 wt% hydrogen peroxide. Aqueous gel permeation chromatography revealed that all the emulsifiers were polymers of high relative molecular mass, M_r (number-average M_r in excess of 125,000). Kinetic and sequence distribution studies, performed using ¹H and ¹³C NMR, revealed complete conversion of all double bonds with the grafts statistically dispersed along the backbone.

Received 2 January; accepted 10 February 1998.

1. Piirma, I. *Polymeric Surfactants* (Vol. 2, Surfactant Science Ser., Dekker, New York, 1992).
2. Laschewsky, A. Molecular concepts, self-organization and properties of polysoaps. *Adv. Polym. Sci.* **124**, 1–86 (1995).

3. Tsuchida, E. & Abe, K. Interactions between macromolecules in solution and intermolecular complexes. *Adv. Polym. Sci.* **45**, 1–119 (1982).
4. Bekturov, E. A. & Bimendina, L. A. Interpolymer complexes. *Adv. Polym. Sci.* **43**, 100–147 (1980).
5. Osada, Y. Conversion of chemical into mechanical energy by synthetic polymers. *Adv. Polym. Sci.* **82**, 1–46 (1987).
6. Kabanov, V. A. & Papisov, I. M. Formation of complexes between complementary synthetic polymers and oligomers in dilute solution. Review. *Vysokomol. Soyed.* **A21**, 243–281 (1979).
7. Scranton, A. B., Klier, J. & Aronson, C. L. in *Polyelectrolyte Gels* (eds Harland, R. S. & Prud'homme, R. K.) 171–189 (ACS Symp. Ser. 480, Am. Chem. Soc., Washington DC, 1992).
8. Klier, J., Scranton, A. B. & Peppas, N. A. Self-associating networks of poly(methacrylic acid-g-ethylene glycol). *Macromolecules* **23**, 4944–4949 (1990).
9. Scranton, A. B., Klier, J. & Peppas, N. A. Complexation thermodynamics of free and graft oligomers with complementary polymers. *J. Polym. Sci.: Polym. Phys.* **29**, 211–224 (1991).
10. Saito, S. & Sakamoto, T. Solubilization by polymer/polymer complexes. *Colloids Surf.* **23**, 99–104 (1987).
11. Hemker, D. J., Garza, V. & Frank, C. W. Complexation of poly(acrylic acid) and poly(methacrylic acid) with pyrene-end-labeled poly(ethylene glycol). pH and fluorescence measurements. *Macromolecules* **23**, 4411–4418 (1990).
12. Noolandi, J. Multiblock copolymers as polymeric surfactants: are “pancakes” better than “dumbbells”? *Makromol. Chem. Theory Simul.* **1**(5), 295–298 (1992).
13. Mayo, S. L., Olafson, B. D. & Goddard, W. A. DREIDING: A generic force field for molecular simulations. *J. Phys. Chem.* **94**, 8897–8909 (1990).

Correspondence and requests for materials should be addressed to A.B.S. (e-mail: scranton@egr.msu.edu).

Global seasonal rainfall forecasts using a coupled ocean–atmosphere model

T. N. Stockdale, D. L. T. Anderson, J. O. S. Alves & M. A. Balmaseda

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK

One conceptual model of weather is that of a series of events which are unconnected. That is, that the weather next week is essentially independent of the weather this week. However, although individual weather systems might be chaotic and unpredictable beyond a week or so, the statistics describing them may be perturbed in a deterministic and predictable way¹, particularly by the ocean. In the past, seasonal forecasts of atmospheric variables have largely been based on empirical relationships, which are weak in most areas of the world². More recently, atmosphere models forced by assumed or predicted ocean conditions have been used^{3,4}. Here a fully coupled global ocean–atmosphere general circulation model is used to make seasonal forecasts of the climate system with a lead time of up to 6 months. Such a model should be able to simulate the predictable perturbations of seasonal climate, but to extract these from the chaotic weather requires an ensemble of model integrations, and hence considerable computer resources. Reliable verification of probabilistic forecasts is difficult, but the results obtained so far, when compared to observations, are encouraging for the prospects for seasonal forecasting. Rainfall predictions for 1997 and the first half of 1998 show a marked increase in the spatial extent of statistically significant anomalies during the present El Niño, and include strong signals over Europe.

The year 1997 saw many reports of anomalous weather conditions in different parts of the globe. Tropical examples were the droughts over Indonesia and Papua New Guinea and parts of Australia, the floods in Ecuador and Peru, and the prolonged torrential rains over East Africa. Extratropical examples were the severe floods in Poland and East Germany in July and the wettest June since 1860 in England, the storms in Hong Kong which produced the heaviest rain since records there began, the worst-ever drought in northern China, and the weakest Atlantic hurricane season for 50 years. There are many other examples. Some are easily linked with the large El Niño currently in progress; for others the situation is less clear.

A simple model of weather is an unbiased coin. If such a coin is

tossed several hundred times, one would expect to find short runs of one face or the other, say five ‘heads’ in a row, purely by chance. If the heads are thought of as inactive weather systems, then such a run of heads might correspond to drought conditions. Such a drought would be unpredictable, as it is simply the outcome of a series of chance processes. But if the coin were biased, the statistics would also be biased, although the outcome of any individual toss of the coin would remain unpredictable. Sea surface temperature (SST) changes, particularly those associated with El Niño in the tropical Pacific, can influence the atmosphere both locally and at great distances, acting to bias the coin in this analogy.

A strategy for seasonal forecasting is to build a model system which includes the ocean and the atmosphere, and to pay particular attention to the ability of the model to predict both El Niño variability and its effects worldwide. These two aspects of seasonal forecasting are often separated^{4,5}, but here we present results from a comprehensive coupled atmosphere–ocean model which combines both in a single stage. The potential advantages of a fully coupled one-stage approach are that the full set of initial conditions (ocean, atmosphere and land surface) are able to influence the evolution of the coupled system, and that a wide range of interactions between ocean and atmosphere is represented during the forecast. General reviews of seasonal forecasting are available elsewhere^{6–8}.

The atmosphere model component has a resolution of 1.875° with 31 vertical levels. The ocean resolution is comparable but in the tropics increases to 0.5° in the latitudinal direction, so as to resolve the equatorial waves which are important for El Niño. The atmospheric and land surface initial conditions are taken from the operational analysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). Ocean initial conditions are taken from an analysis of the ocean state made using the forced response of the ocean, and assimilating surface temperature analyses and all available sub-surface thermal ocean data^{9,10}.

The forecast strategy is to couple the atmosphere and ocean directly without flux corrections and integrate forward for 6 months from the initial conditions. This inevitably leads to a drift in the climate of the coupled system, due to systematic errors in the component models. Such biases, which are sometimes large, can be substantially corrected—after a model integration is complete—by removing the drift as estimated from a set of forecasts for an earlier period¹¹. Typically, the SST drift in the equatorial Pacific is 1–2 °C over a 6-month period. The drift has been removed in all the plots shown here. Extensive experimentation for the 1980s and especially the 1990s shows that our system has useful predictive skill. The forecasts that we present here come from our quasi-operational system. Three forecasts are made from each week’s ocean analysis with slightly different atmospheric initial conditions to increase our sampling of the chaotic part of the system.

Figure 1a shows the observed SST anomalies in December 1996, when a weak cold anomaly was present in the east Pacific. Our 6-month forecast anomaly for May, created by averaging the forecasts starting at different days in December, is shown in Fig. 1b. This ensemble mean forecast is quite similar to that which was later observed to occur, Fig. 1c. This period was chosen as it shows that the onset was well predicted in our 6-month lead forecasts. Further illustration of the temporal consistency of our forecasts is shown in Fig. 2. Not only was the onset well predicted, but subsequent forecasts have also represented the rapid growth rather well. Although the amplitude of the event is underestimated in the longer range of some of the forecasts, in general the predictions of the large amplitude have been good.

There can be appreciable differences between individual forecasts, as illustrated in Fig. 2. Three forecasts initiated for 21, 22 and 23 December 1996 are shown by heavy dashed lines: the ocean state is the same in all three, and only the initial atmospheric state changes. All show the overall warming, but the size of the warming differs by up to 0.5 °C. Other examples can be found where the

spread is larger. This variability indicates a fundamental limit to predictive skill, as the differences arise simply from different synoptic atmospheric events which are unpredictable on these timescales^{12–15}. There are other factors which at present limit predictive skill: the model is not perfect, and the ocean's initial conditions are not known accurately. With better models and better observing systems, these limitations could be reduced, but a non-trivial uncertainty in the evolution of SST appears to be unavoidable.

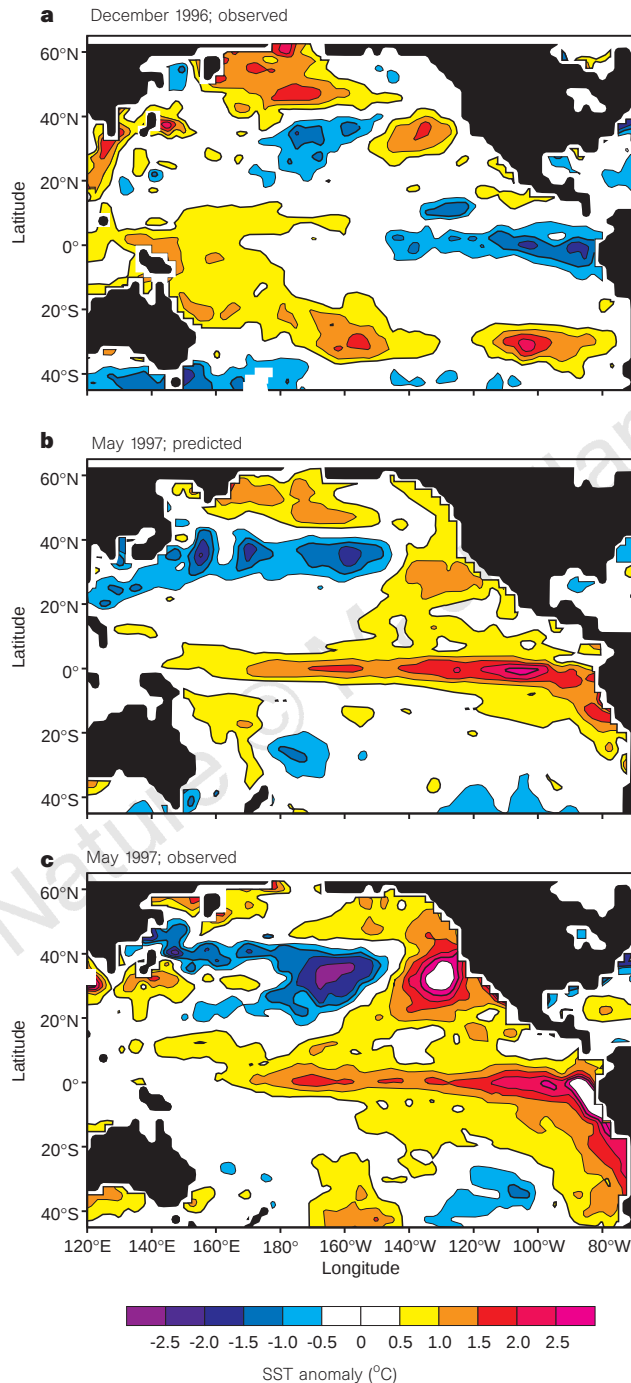


Figure 1 Plot of SST anomalies in the Pacific sector. **a**, Observed anomaly in December 1996; **b**, predicted SST anomaly for May 1997 from forecasts initiated during December 1996; **c**, observed anomaly in May 1997. Contour interval is 0.5°C. The change in equatorial SST between December and May was well predicted.

Although prediction of tropical SST is a basic requirement for seasonal forecasts, or greater practical importance is whether we can predict quantities such as rainfall. Our approach is direct: we take the atmosphere model output directly from our 'single-stage' coupled forecast. Rainfall shows a substantially larger spread between individual forecasts than does SST, and this can make it difficult to establish unambiguously whether an ensemble-mean rainfall pattern is meaningfully different from average conditions. By taking sufficiently large ensembles and by having appropriate climatological control samples, statistical tests can be used to assess whether an apparent signal is in fact significant.

In Fig. 3a we show the ensemble-mean rainfall anomalies for June–August 1997, as predicted by our coupled system from 27 initial dates in March and April. The reference climate is made up of 114 members, each of which is a 6-month run of the coupled model from 1 April; 19 integrations for each of the years 1991–96. This is not a true long-term climate, but should approach it over most of the Earth. For each point a Wilcoxon–Mann–Whitney test (see, for example, ref. 16) is applied to distinguish whether the ensemble population for 1997 is shifted relative to that of the reference climate. This test is used because it is not sensitive to the nature of the distributions: rainfall distributions can be quite non-normal. We only plot the points which are significant at the 99% level. The Bayesian probability that a given shaded area is due to a real signal and not chance depends on the actual amount of shading in the plot. The choice of a 99% local significance level is a conservative compromise. Most of the shaded areas in these plots are likely to be due to real signals, while many areas which are likely to have a real signal remain unshaded.

Figure 3a shows a relatively high level of signal in the tropics, and several rainfall features often associated with El Niño. This might be expected, given that the coupled model correctly forecast the onset of El Niño. Both India and Eastern Australia have a dry signal in the forecasts, but it is too weak to pass our conservative significance test. (As it turned out, the Indian monsoon was close to normal in 1997 despite El Niño). The most striking mid-latitude feature is the wet signal over south and central Europe. This is highly significant (the level is >99.999% for points in the central regions), and more than 80% of the ensemble members give above-average rainfall. Such

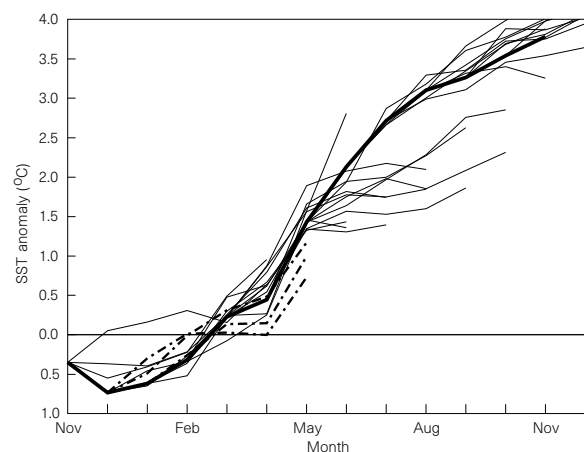


Figure 2 Plume of monthly mean SST anomalies predicted for the Niño-3 region (5°S–5°N, 90°–150°W). Forecasts are initiated at weekly intervals from November 1996 until April 1997, and run for 6 months. The heavy line shows the observed values. The onset in May was well captured by the forecasts, as was the large amplitude of the event. The heavy dashed lines show three ensemble members for atmospheric initial conditions 21, 22, 23 December for the same ocean initial conditions. The spread, originating from chaotic atmospheric processes, indicates a limit to predictability.

strong and consistent signals over Europe were unexpected, and on the basis of ensemble integrations for the previous 6 years are unusual.

Proper verification of our seasonal forecasts would require large ensembles to define the forecast probability distributions, and large numbers of independent forecast periods to test statistically the relationship between forecast probability and actual outcome. For any individual case, one can never be sure that a match between forecast and observations is not simply fortuitous. Nonetheless, the unusual strength of the wet signal over Europe and its close

matching in space and time to observed exceptional wet weather leads us to conclude that it is likely that there is a physical connection. If true, this means that the wet European summer weather is not in the category of a purely chance event, as discussed earlier, but rather is related to a predictable cause. It is tempting to link the rain with El Niño. This is a reasonable hypothesis, and is strengthened by our demonstration of a physical cause of the rainfall. Correlation is not the same as causation, however, and controlled experiments would be needed to confirm the link as there are other potential sources of predictability. Further evidence for

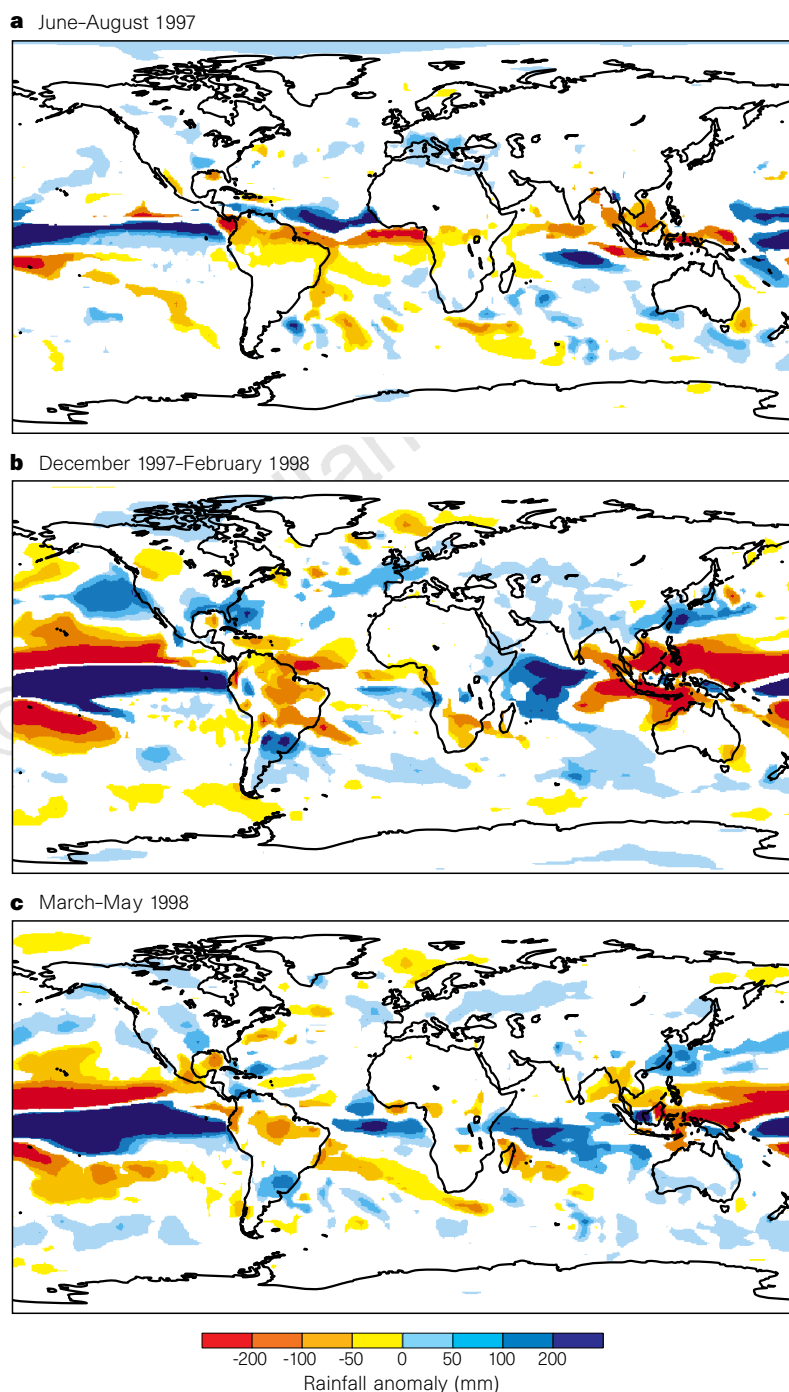


Figure 3 Global plots of predicted rainfall anomalies. **a**, For June–August 1997 from forecasts initiated in March and April; **b**, for December 1997 to February 1998 from forecasts initiated in September and October; **c**, March–May for 1998 from forecasts initiated in December and January. Only statistically significant (>99%)

anomalies are shown. Particularly unusual in **a** is the anomaly over south/central Europe. **b** and **c** show that significant precipitation signals are very extensive during this El Niño event, and are not restricted to the tropics.

physical causation of the excess rainfall comes from ensemble integrations of the uncoupled atmosphere model forced by observed SST, which also produce a strong wet signal over Europe for summer 1997.

The ensemble mean precipitation anomalies forecast for December 1997 to February 1998 are shown in Fig. 3b. These forecasts start in September and October, and in all cases the predicted SST shows El Niño conditions at least as strong as 1982–83 during northern winter. The plot shows an even stronger level of signal than summer 1997, and the general pattern of the precipitation anomalies is very plausible for a strong El Niño event. We note, for example, the wet signal in California, Florida, Uruguay and East Africa, and the dry signal in the Amazon region and southern Africa. More interesting are those features which might not have been expected empirically, but have apparently occurred: wet conditions in India and southeast China, a lack of drought in northern Australia, and wet weather in western Europe. Figure 3c shows our latest forecast, for the period March–May 1998. The model rainfall distribution is again significantly shifted in many regions of the world.

Empirical studies show evidence of links between some aspects of European weather and El Niño and other factors, but correlations are relatively low^{2,17} and an El Niño signal tends to be visible only in winter months. As far as we know, no empirical scheme would have predicted the summer 1997 European rainfall from the data available in April. One should not draw too much from a single case, but it seems that at least in some instances a model-based approach goes beyond what present-day empiricism can do.

It is our belief that dynamical methods offer the best long-term hope for making seasonal forecasts because of their greater generality and precision: that is, because of their ability to handle unprecedented situations and to treat nonlinear combinations of factors which cannot be extracted empirically from the short observational records available to us. Dynamical models still have relatively large errors, and cannot yet be considered reliable. Nonetheless, our results suggest that useful model-based seasonal forecasting is now possible. Future improvements in models and their initialization can only increase the reliability and usefulness of such forecasts in the years to come. □

Received 21 August 1997; accepted 25 February 1998.

1. Palmer, T. N. Extended range atmospheric prediction and the Lorenz model. *Bull. Am. Meteorol. Soc.* **74**, 49–65 (1993).
2. Barnston, A. G. Linear statistical short-term climate predictive skill in the northern hemisphere. *J. Clim.* **7**, 1513–1564 (1994).
3. Barnett, T. P. et al. Forecasting global ENSO-related climate anomalies. *Tellus A* **46**, 367–380 (1994).
4. Livezey, R. E., Masutani, M. & Ji, M. SST-forced seasonal simulation and prediction skill for versions of the NCEP/MRF model. *Bull. Am. Meteorol. Soc.* **77**, 507–517 (1996).
5. Ji, M., Leetmaa, A. & Koussy, V. E. Coupled model predictions of ENSO during the 1980s and the 1990s at the National Centers for Environmental Prediction. *J. Clim.* **9**, 3105–3120 (1996).
6. Palmer, T. N. & Anderson, D. L. T. The prospects for seasonal forecasting: a review. *Q. J. R. Meteorol. Soc.* **120**, 755–793 (1995).
7. Hastenrath, S. Recent advances in tropical climate prediction. *J. Clim.* **8**, 1519–1532 (1995).
8. Livezey, R. E. et al. Teleconnective response of the Pacific-North American region atmosphere to large central Pacific SST anomalies. *J. Clim.* **10**, 1787–1820 (1997).
9. Reynolds, R. W. & Smith, T. M. Improved global sea surface temperature analyses using optimum interpolation. *J. Clim.* **7**, 929–948 (1994).
10. Smith, N., Blomley, J. E. & Meyers, G. A univariate statistical interpolation scheme for subsurface thermal analyses in the tropical oceans. *Prog. Oceanogr.* **28**, 219–256 (1991).
11. Stockdale, T. N. Coupled ocean-atmosphere forecasts in the presence of climate drift. *Mon. Weath. Rev.* **125**, 809–818 (1997).
12. Blanke, B., Neelin, J. D. & Gutzler, D. Estimating the effect of stochastic wind stress forcing on ENSO irregularity. *J. Clim.* **10**, 1473–1486 (1997).
13. Eckert, C. & Latif, M. Predictability of a stochastically forced hybrid coupled model of El Niño. *J. Clim.* **10**, 1488–1504 (1997).
14. Kleeman, R. & Moore, A. M. A theory for the limitation of ENSO predictability due to stochastic atmospheric transients. *J. Atmos. Sci.* **54**, 753–767 (1997).
15. Stockdale, T., Latif, M., Burgers, G. & Wolff, J.-O. Some sensitivities of a coupled ocean-atmosphere GCM. *Tellus A* **46**, 367–380 (1994).
16. Wonnacott, T. H. & Wonnacott, R. J. *Introductory Statistics* (Wiley, New York, 1977).
17. Fraedrich, K. An ENSO impact on Europe? A review. *Tellus A* **46**, 541–552 (1994).

Acknowledgements. The model integrations were carried out on a VPP300 machine donated by Fujitsu Limited. The ocean model was provided by the Max-Planck-Institut für Meteorologie, Hamburg; ocean data assimilation software by the Bureau of Meteorology Research Centre, Melbourne; and coupling software by CERFACS, Toulouse.

Correspondence and requests for materials should be addressed to T.N.S. (e-mail: t.stockdale@ecmwf.int). Further information and forecast plots can be found on the ECMWF Seasonal Forecast Project Web pages at <http://www.ecmwf.int>

Ice-sheet variability around the North Atlantic Ocean during the last deglaciation

A. Marshall McCabe* & Peter U. Clark†

* School of Environmental Studies, University of Ulster, Coleraine, County Londonderry, Northern Ireland BT52 1SA, UK

† Department of Geosciences, Oregon State University, Corvallis, Oregon 97331, USA

Millennial-scale variability in the flux of ice-rafted detritus to North Atlantic sediments during the last glacial period has been interpreted to reflect a climate-forced increase in the discharge of icebergs from ice-sheet margins surrounding the northern North Atlantic Ocean¹. But the relationship between ice-sheet variability and climate change is not clear, as both the sources of ice-rafted detritus and the ice-marginal processes are varied and complex^{2–4}. Terrestrial records are helpful in unravelling this complexity because they can demonstrate the scale of ice-sheet oscillations, and whether the ice sheet (or sector) was advancing or retreating with respect to climate change. Here we constrain the age and anatomy of a prominent readvance of the British Ice Sheet in the northern Irish Sea region at ~14 ¹⁴C kyr BP (~16.4 calendar kyr BP). The analysis indicates that the British Ice Sheet participated in an iceberg discharge episode known as Heinrich event 1. Comparison with other terrestrial and marine ice-sheet records suggests that the dynamic collapse of the Laurentide Ice Sheet beginning at 14.6–15.0 ¹⁴C kyr BP^{1,4} (~17.2–17.6 calendar kyr BP)⁵ initiated varied responses from other ice-sheet margins around the northern North Atlantic region. These observations support the argument that the release of icebergs and meltwater during Heinrich event 1 disrupted the North Atlantic thermohaline circulation^{6–8}, leading to a delay or reversal of deglaciation of the Northern Hemisphere and at least as far south as 40° S for two to three thousand years^{9,10}, suggesting a climate forcing and response similar to that of the ensuing Younger Dryas ‘cold snap’^{11,12}.

Terrestrial records of millennial-scale ice-sheet oscillations should be present in the climatically sensitive northeastern Atlantic region, but they have a low preservation potential along continental margins, reflecting the duration and erosional effects of ice cover on the continental shelf. However, geological records of oscillations of the Irish Sea basin ice stream, which was a major conduit for British Ice Sheet (BIS) drainage during the last deglaciation (Fig. 1), are well preserved because they offlap northwards (inland) towards centres of ice dispersion. This ice stream experienced at least five ice-marginal oscillations between 22 and 14 ¹⁴C kyr BP¹³, suggesting similar millennial-scale variability as identified for other amphih-North Atlantic ice sheets. The phasing relations between BIS variability and that of other ice sheets, however, has not been rigorously established.

On the northwestern margins of the Irish Sea basin, the outer limit (130 km long) of a major southeasterly ice readvance from the Irish lowlands is marked by a narrow (1–3 km) terminal zone of ice-contact landforms (Fig. 1). On the margins of Dundalk Bay and Carlingford Lough, moraines and deltas overlies interstadial muds and mark the former extent of grounded ice lobes that ended in shallow marine settings. To the northeast at Killard Point, the same ice limit is marked by a raised ice-contact morainal bank which prograded southeastwards into the sea. The bank is dominated by boulder gravel and marine mud within multistoried channels which record a high sediment flux to the ice margin. The presence of raised, late-glacial shoreface gravels immediately beyond these limits indicates synchronicity between ice lobe maxima and high relative sea level.

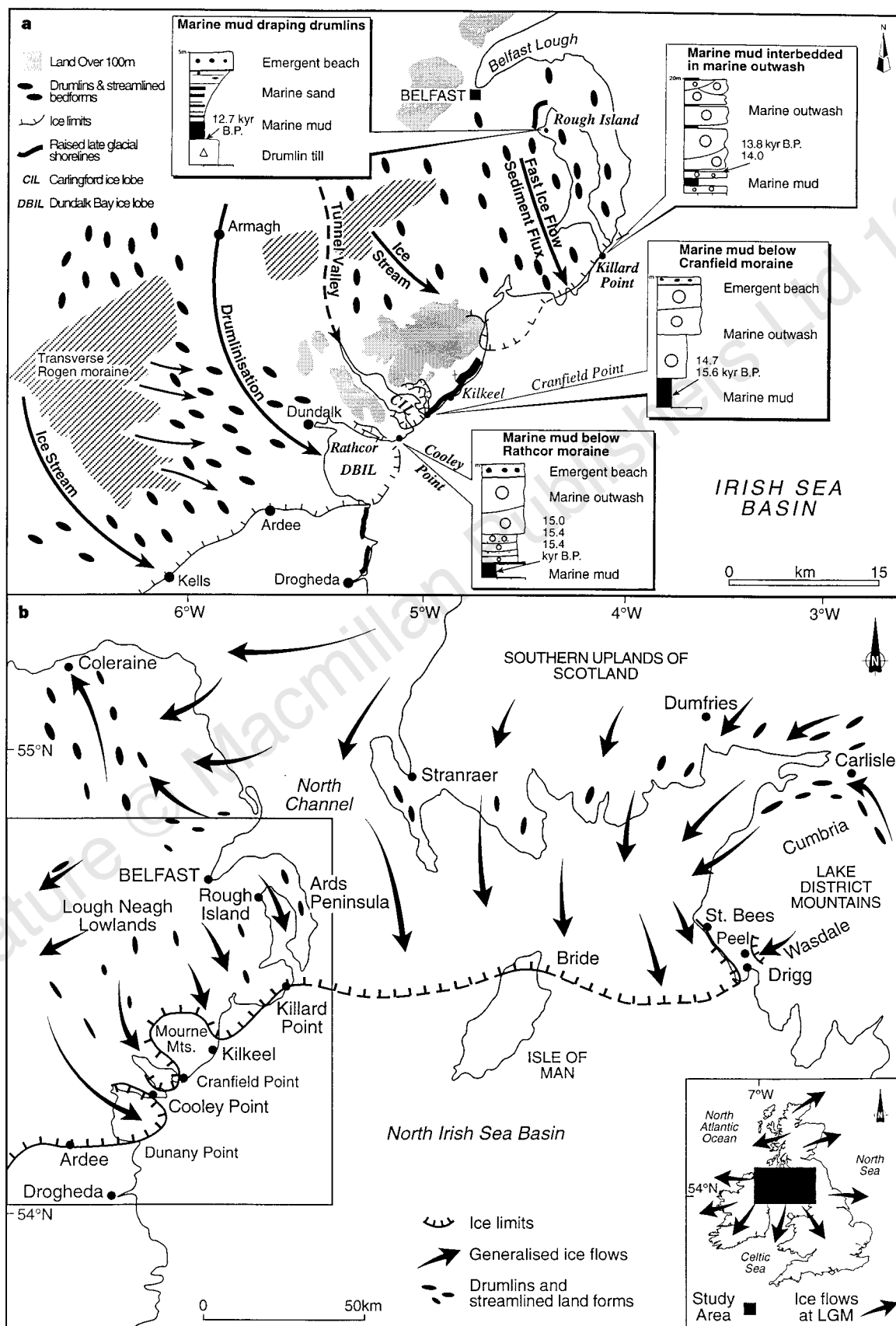


Figure 1 a, Deglacial records in the north Irish Sea basin. Pattern of ice flow during H1 and the stratigraphic position of dated marine mud beds in the north

Irish Sea basin. **b**, Ice flows and H1 ice limits in the north Irish Sea basin. Inset, ice flows in Britain during the LGM.

Drumlins occur within 1–6 km up ice of the terminal moraines. Ice flow lines identified from these bedforms indicate contemporaneity of drumlinization (fast ice flow) and moraine building. Inland from Dundalk, drumlinization along a curved flow path crosscuts and overprints a prominent zone of Rogen moraine ridges, suggesting that bed reorganization and sediment transfer to ice margins occurred during ice readvance (Fig. 1). The belt of pristine Rogen moraine 25 km inland from Dundalk suggests that the ice sheet had only retreated a short distance before drumlinization and ice readvance. The presence of a regional drape of red marine mud directly overlying drumlin tills in the Ards Peninsula, and the absence of fluvio-glacial deposits inland of the moraines, indicates that drumlinization was followed closely by rapid ice retreat.

On the northern margins of the Irish Sea basin, regional ice flow lines reconstructed from drumlins are also bounded by limiting moraines on the Isle of Man (Bride) and along the Cumbrian coast (St Bees, Drigg). The pattern of ice flow associated with these moraines is similar to the readvance across eastern Ireland and suggests a regional pattern of convergent ice flow from separate ice centres into the northern Irish Sea basin at this time (Fig. 1).

The age of this readvance is constrained by accelerator mass spectrometry ^{14}C dating of the benthonic foraminifera *Elphidium clavatum* sampled from marine muds below and interbedded with ice-contact outwash at four sites along the ice limit (Fig. 1 and Table 1). Microfaunas in these mud beds are dominated by *E. clavatum* (85–95%) and the ostracod *Roundstonia globulifera* (5–10%) which shows intact instars. Similar opportunistic biocoenoses have been recorded from contemporary Arctic–subarctic areas recently vacated by tidewater glaciers, and are typical of glaciomarine environments (0.5–2.5 °C, turbid, lowered salinities)¹⁴. The AMS ^{14}C dates (Table 1) and lithostratigraphy (Fig. 1) provide a coherent time frame for palaeoenvironmental changes occurring before, during and after the readvance in the north Irish Sea basin (Cooley Point interstadial, from ≥ 16.7 and ≤ 14.7 ^{14}C kyr BP; Killard Point stadial, with ice margin advance after 14.7 ^{14}C kyr BP and at its maximum position at 14.0 ^{14}C kyr BP; Rough Island interstadial with rapid retreat after 13.8 ^{14}C kyr BP). Our results from these terrestrial records thus show that rapid ice accumulation and advance in the northern sectors of the BIS occurred during Heinrich event 1 (H1) following early widespread deglaciation along the southern part of the ice sheet between ≥ 16.7 and ≤ 14.7 ^{14}C kyr BP. The Killard Point Stadial therefore documents a major deglacial oscillation of the BIS preceding the Loch Lomond stadial, which is marked by renewed ice-sheet development in the highlands of western Scotland during the Younger Dryas¹⁵ (10–11 ^{14}C kyr BP).

The new data from the Irish Sea basin provide critical new information for comparing ice-marginal events in the amphinorth Atlantic during H1. In Fig. 2, we compare terrestrial and marine records which demonstrate that several sectors of several ice sheets around the North Atlantic advanced during H1, whereas other ice sheets began to retreat at or shortly following the beginning of H1. Ice-rafted detritus (IRD) records suggest synchronous increases in the flux of icebergs delivered from ice streams draining the Laurentide Ice Sheet (LIS) through Hudson Strait and the Gulf of St Lawrence beginning 14.6 to 15 ^{14}C kyr BP¹⁴ (Fig. 2d). Records from the southern margin of the LIS suggest that, following a major retreat that ended at 15.0–15.5 ^{14}C kyr BP (Erie interstade), the margin readvanced (Port Bruce stade, Crown Point phase) several hundred kilometres to within <150 km of its maximum Late Wisconsin position and then retreated rapidly during or near the end of H1 (refs 16, 17; Fig. 2e). Our data from the northern Irish Sea area suggest a similar sequence of events for the BIS, with an interstadial preceding a readvance of the ice margin which was at its maximum position at 14 ^{14}C kyr BP, and was followed by rapid retreat after 13.8 ^{14}C kyr BP (Fig. 2c). We do not know precisely when the BIS began to readvance, but our data suggest that it was at its

maximum position at a time (13.8–14 ^{14}C kyr BP) when retreat of the southern margin of the LIS was well underway (Fig. 2e). Bond and Lotti¹ argue from IRD data that the Icelandic Ice Sheet (IIS) began to discharge icebergs at the same time as marine ice streams that drained the LIS. The eastern Greenland Ice Sheet (GIS) began to

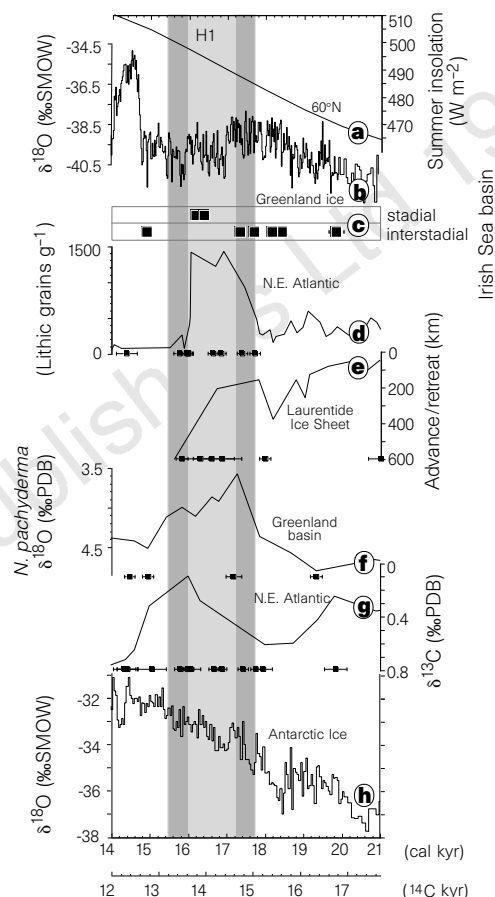


Figure 2 Records spanning the interval from 14–21 calendar kyr (12–17.7 ^{14}C kyr BP, using the calibration of ref. 5) showing varied responses of the climate system at Heinrich event 1 (H1). Uncertainties in the age of the beginning (14.6–15.0 ^{14}C kyr BP) and end (13.2–13.6 ^{14}C kyr BP)¹⁴ of H1 are shown by the vertical grey bars. **a**, Insolation values for June at 60° N (in W m^{-2}). **b**, GRIP ice core $\delta^{18}\text{O}$ record (20-yr means from 14–20 calendar kyr, 50-yr means from 20–21 calendar kyr)²⁸. Note that the cooling indicated by the $\delta^{18}\text{O}$ record (for example, Oldest Dryas) occurred after the beginning of H1 in the GRIP chronology, whereas it occurred at the beginning of H1 in the GISP2 chronology (not shown), a chronology issue. **c**, Radiocarbon ages on our samples from the Irish Sea basin (Table 1) (errors shown where larger than the size of the symbol). Ages constraining glacial advance versus retreat are shown as 'stadial' versus 'interstadial', respectively. **d**, Record of concentration of lithic grains per gram of sediment from North Atlantic core VM23-081 (ref. 1) (54° 2' N, 16° 8' W) showing rise in lithic grains during H1. Radiocarbon ages (with errors) constraining the chronology of this record²⁹ are shown along the base. **e**, Record of advance and retreat (in km) of the Lake Michigan lobe of the LIS¹⁷. Ice-margin advance is directed up-page; 0 km point is arbitrarily located. Radiocarbon ages (with errors) constraining the age of ice-margin advance during the Crown Point phase¹⁷ are also shown along the base of the record. **f**, $\delta^{18}\text{O}$ record (in ‰) from core HM94-23 measured on *N. pachyderma* (s.) from the Greenland basin²⁹ (73° 46' N, 02° 23' W). Radiocarbon ages (with errors) constraining the chronology of this record²⁹ are shown along the base. The large $\delta^{18}\text{O}$ anomaly (more negative values) reflects the disintegration of the Barents Sea (and possibly the Greenland and Fennoscandian) ice sheet(s) during H1. **g**, $\delta^{13}\text{C}$ record (in ‰) *C. wuellerstorfi* from core VM23-81 (ref. 6). Age model is based on radiocarbon ages²⁰ which are shown (with errors) along the base of the record. More negative values of $\delta^{13}\text{C}$ during H1 indicate a reduction in the strength of NADW. **h**, Byrd, Antarctica, $\delta^{18}\text{O}$ ice core record (in ‰) ($\delta^{18}\text{O}$ values from ref. 30; age model of ice core from ref. 21).

Table 1 AMS ^{14}C ages from the Irish Sea basin

Locality/Description/Comments	Laboratory number	$\delta^{13}\text{C}$	Radiocarbon age* (yr BP $\pm 1\sigma$)
Rough Island: Samples from marine mud draping drumlins. Dates marine transgression after ice sheet collapse	AA21822	-2.089	12,740 $\pm$ 95
Killard Point: Samples from marine mud interbedded in outwash from the Killard Point moraine. Dates ice sheet readvance	AA22820	-2.607	13,795 $\pm$ 115
	AA22821	-2.702	13,955 $\pm$ 105
Cranfield Point: Samples from interstadial mud below outwash from the Carlingford Bay ice lobe. Dates ice free interval before readvance	AA21818	-2.097	14,705 $\pm$ 130
	AA21819	-1.407	15,605 $\pm$ 140
Cooley Point: Samples from interstadial mud below outwash from the Dundalk Bay ice lobe. Records ice free interval before readvance	AA17693	-2.4	15,020 $\pm$ 110
	AA17694	-2.5	15,390 $\pm$ 110
	AA17695	-2.5	15,400 $\pm$ 140
Kilkeel: Samples from interstadial mud below outwash from the Carlingford Bay ice lobe. Dates ice free interval before readvance	AA22351	-2.019	16,760 $\pm$ 130
	AA22352	-2.375	16,750 $\pm$ 160

* AMS ^{14}C ages from monospecific samples of *Elphidium clavatum* in marine muds. Samples were submitted to the NSF-AMS facility at the University of Arizona as CO_2 . Samples corrected for assumed 400-yr difference between surface-water carbon and atmospheric carbon. Sampling localities shown in Fig. 2.

discharge IRD-laden icebergs between 14.5 and 15.3 ^{14}C kyr BP¹⁸. Finally, marine-based margins of the Barents Sea and Fennoscandian Ice Sheets (BSIS and FIS) began to retreat 14.7–15.0 ^{14}C kyr BP from their maximum positions which they had maintained since the last glacial maximum (Fig. 2f)^{11,19} with icebergs discharged from the retreating margins delivering IRD to the adjacent continental slope and Norwegian Sea¹⁹.

The varied behaviour of amphi-North Atlantic ice sheets during H1 has significant implications for interpreting IRD signals because of the assumed link between ice-sheet advances, which would have delivered IRD to marginal seas, and climate forcing. Comparison of IRD records to the GRIP and GISP2 ice-core records suggests that most Heinrich events and the higher-frequency IRD events occurred at the end of cooling intervals and were followed by abrupt warming (interstadials)^{1,20}. Bond and Lotti¹ argued that these events were triggered by climate or a climate-related mechanism because of the relation of the IRD events to cool (stadial) climates, and the fact that more than one ice sheet was involved in discharging icebergs and IRD into the North Atlantic. The record of retreat of the FIS¹¹ and BSIS¹⁹ during H1 (Fig. 2f), however, indicates that an increase in the flux of IRD-charged icebergs to the open ocean is not necessarily related to a climatically forced ice-margin advance. Insofar as the IRD record of H1 is comparable to deep-sea records of other IRD events, the varied response of amphi-North Atlantic ice margins during H1 suggests the potential for more complex interpretations of other IRD events.

A further implication of our results suggests that H1 is an exception to the common relation between stadial climates and IRD events²⁰ because it reversed a warming trend clearly seen in Greenland ice-core records (Fig. 2b) and other North Atlantic records⁹ that began about 21 calendar kyr BP, when summer insolation at 60°N began to increase (Fig. 2a). Cooling occurred during H1 (for example, Oldest Dryas) as a result of disruption of North Atlantic Deep Water (NADW) formation (and its associated heat transport) by release of meltwater and icebergs (Fig. 2g)^{6–8} and continued until the abrupt warming at the onset of the Bølling–Allerød warm interval. Greenland ice cores suggest that temperatures at the onset of the Bølling–Allerød reached a level (near interglacial) that was in line with the warming trend started before H1 (Fig. 2b). The warming trend that began in the North Atlantic region about 21 calendar kyr BP also began in Antarctica (Fig. 2h). However, during H1 this warming trend was interrupted in the Northern Hemisphere and as far south as 40°S^{5,10}, but it continued over most of Antarctica throughout the last deglaciation^{21,22} (Fig. 2h). A similar climatic response is associated with discharge of meltwater and icebergs from the LIS during the Younger Dryas^{11,12}.

In the context of a climate forcing mechanism for Heinrich events¹, the relation of H1 to a warming trend suggests that the cause of H1 may have differed from the cause of the other IRD events²⁰. Alternatively, this relation is consistent with a suggestion²³ that Heinrich events originated by an internal instability of the LIS

that is independent of climate forcing. This mechanism has been questioned¹ because it does not explain either increased iceberg discharges from multiple ice-sheet sectors and multiple ice sheets, or the increased flux of icebergs from ice streams before discharge along Hudson Strait, as interpreted from the IRD record. However, the example of H1 shows that IRD data in general have the potential for representing a varied and complex response of ice-sheet margins during iceberg-discharge events. Furthermore, lithic data on H1 from deep-sea core VM23-081 (ref. 1) suggest that IRD from various sources, including Hudson Strait, increased at the same time (15 ^{14}C kyr BP). Within the constraints of the radiocarbon chronology, the southern LIS margin was also advancing at this time (Fig. 2e), indicating that much of the LIS margin advanced during H1 (ref. 16). Before H1, the southern LIS margin had been in overall retreat (with secondary, asynchronous oscillations) during the warming trend recorded in Greenland ice cores¹⁶. This retreat in response to warming is expected, particularly given the extremely low ice-surface gradients that characterized the southern margin²⁴. It is unlikely that the observed warming trend before H1 initiated an ice-sheet surge, however, because the effect of such a change in temperature at the ice surface will be strongly attenuated with depth, especially through a thick ice sheet like the LIS²³.

We believe that the timings and patterns of ice-sheet variability in the North Atlantic region discussed here (Fig. 2) are most consistent with the argument that an internal instability of the LIS leading to H1 initiated a sequence of events that were ultimately responsible for reversing initial deglacial warming and thus delaying the transition from full glacial to interglacial conditions at least as far south as 40°S. In particular, discharge of icebergs to the North Atlantic from the LIS would have rapidly reduced formation of NADW^{6–8}, resulting in cooling of the North Atlantic region^{9,25} and causing the Oldest Dryas. The apparent synchronous increase in iceberg discharge from the IIS and LIS¹ may reflect the rapid response of a small ice sheet to this cooling. Alternatively, the example from the BSIS IRD signal¹⁹ suggests that the IIS and GIS¹⁸ IRD signals during H1 may correspond to increased iceberg discharge during retreat of their marine margins. IRD data alone cannot distinguish between these two fundamentally different behaviours. Our evidence for the BIS readvance during H1 is consistent with response of an ice sheet in this climatically sensitive region to North Atlantic cooling. Retreat of the FIS and BSIS margins may also reflect a climatic influence through a reduction of precipitation at these latitudes in association with the cooler North Atlantic region²⁵. In addition, their thinned marine margins may have also responded to sea-level forcing associated with the partial collapse of the LIS²⁶. □

Received 15 August 1997; accepted 19 January 1998.

1. Bond, G. & Lotti, R. Iceberg discharges into the North Atlantic on millennial time scales during the last glaciation. *Science* **267**, 1005–1010 (1995).
2. Alley, R. B. & MacAyeal, D. R. Ice-rafted debris associated with binge/purge oscillations of the Laurentide ice sheet. *Paleoceanography* **9**, 503–511 (1994).

3. Gwiazda, R. H., Hemming, S. R. & Broecker, W. S. Tracking the sources of icebergs with lead isotopes: the provenance of ice-rafted debris in Heinrich layer 2. *Paleoceanography* **11**, 77–93 (1996).
4. Jennings, A., Tedesco, K. A., Andrews, J. T. & Kirby, M. E. in *Paleoceanography of the North Atlantic Margins* (eds Andrews, J. T., Austin, W. E. N., Bergsten, H. & Jennings, A. E.) 29–49 (Geol. Soc. spec. publ. no. 111, London, 1996).
5. Bard, E., Rostek, F. & Sonzogni, C. Interhemispheric synchronicity of the last deglaciation inferred from alkenone paleothermometry. *Nature* **385**, 707–710 (1997).
6. Jansen, E. & Veum, T. Evidence for two-step deglaciation and its impact on North Atlantic deep-water circulation. *Nature* **343**, 612–616 (1990).
7. Keigwin, L. D., Jones, G. A., Lehman, S. J. & Boyle, E. A. Deglacial meltwater discharge, North Atlantic deep circulation and abrupt climate change. *J. Geophys. Res.* **96**, 16811–16826 (1991).
8. Keigwin, L. D. & Lehman, S. J. Deep circulation change linked to Heinrich event 1 and Younger Dryas in a mid-depth North Atlantic core. *Paleoceanography* **9**, 185–194 (1994).
9. Bard, E. *et al.* Retreat velocity of the North Atlantic polar front during the last deglaciation determined by ^{14}C accelerator mass spectrometry. *Nature* **328**, 791–794 (1987).
10. Lowell, T. V. *et al.* Interhemispheric correlation of late Pleistocene glacial events. *Science* **269**, 1541–1549 (1995).
11. Lehman, S. J. *et al.* Initiation of Fennoscandian ice-sheet retreat during the last deglaciation. *Nature* **349**, 513–516 (1991).
12. Björck, S. *et al.* Synchronized terrestrial-atmospheric deglacial records around the North Atlantic. *Science* **274**, 1155–1160 (1996).
13. McCabe, A. M. Dating and rhythmicity from the last deglacial cycle in the British Isles. *J. Geol. Soc. Lond.* **153**, 499–502 (1996).
14. Hald, M. *et al.* Recent and Late Quaternary distribution of *Elphidium exilatum* F. *clavatum* in Arctic seas. *Cush. Found. Spec. Publ.* **32**, 141–153 (1994).
15. Sissons, J. B. The Loch Lomond Stadial in the British Isles. *Nature* **280**, 199–203 (1979).
16. Clark, P. U. Unstable behaviour of the Laurentide Ice Sheet over deforming sediment and its implications for climate change. *Quat. Res.* **41**, 19–25 (1994).
17. Hansel, A. K. & Johnson, W. H. Fluctuations of the Lake Michigan lobe during the late Wisconsin subepisode. *Sver. Geol. Unders.* **81**, 133–144 (1992).
18. Stein, R., Nam, S.-I., Grobe, H. & Hubberten, H. in *Paleoceanography of the North Atlantic Margins* (eds Andrews, J. T., Austin, W. E. N., Bergsten, H. & Jennings, A. E.) 135–151 (Geol. Soc. spec. publ. no. 111, London, 1996).
19. Elverhøi, A. *et al.* The growth and decay of the late Weichselian ice sheet in western Svalbard and adjacent areas based on provenance studies of marine sediments. *Quat. Res.* **44**, 303–316 (1995).
20. Bond, G. *et al.* Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
21. Sowers, T. & Bender, M. Climate records covering the last deglaciation. *Science* **269**, 210–214 (1995).
22. Sucher, C. *et al.* Climate records from the Taylor Dome ice core during the last glacial termination. *Eos* **77**, F429 (1996).
23. MacAyeal, D. R. Growth/purge oscillations of the Laurentide ice sheet as a cause of the North Atlantic's Heinrich events. *Paleoceanography* **8**, 775–784 (1993).
24. Clark, P. U. Surface form of the southern Laurentide Ice Sheet and its implications to ice sheet dynamics. *Geol. Soc. Am. Bull.* **104**, 595–605 (1992).
25. Fawcett, P. J., Agustsdóttir, A. M., Alley, R. B. & Shuman, C. A. The Younger Dryas termination and North Atlantic Deep Water formation: Insights from climate model simulations and Greenland ice cores. *Paleoceanography* **12**, 23–38 (1997).
26. Locker, S. D., Hine, A. C., Tedesco, L. P. & Shinn, E. A. Magnitude and timing of episodic sea-level rise during the last deglaciation. *Geology* **24**, 827–830 (1996).
27. Berger, A. & Loutre, M. F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10**, 293–317 (1991).
28. Johnsen, S. J. *et al.* Irregular glacial interstadials in a new Greenland ice core. *Nature* **359**, 311–313 (1992).
29. Koc, N. & Jansen, E. Response of the high-latitude northern hemisphere to orbital climate forcing: Evidence from the Nordic Seas. *Geology* **22**, 523–526 (1994).
30. Johnsen, S. J., Dansgaard, W., Clausen, H. B. & Langway, C. C. Oxygen isotope profiles through the Antarctic and Greenland ice sheets. *Nature* **235**, 429–434 (1972).

Acknowledgements. We thank S. Johnsen for the GRIP data, T. Sowers for the Byrd data, D. Bowen and A. Mix for discussion, and R. Alley and J. T. Andrews for reviewing the manuscript. This work was supported by the NSF (P.U.C.).

Correspondence and requests for materials should be addressed to A.M.M.C. (e-mail: m.mccabe@ulst.break.ac.uk).

Low-sea-level emplacement of a very large Late Pleistocene 'megaturbidite' in the western Mediterranean Sea

R. G. Rothwell, J. Thomson & G. Kähler

Southampton Oceanography Centre, European Way, Empress Dock, Southampton SO14 3ZH, UK

Large-volume turbidites, termed 'megaturbidites' or 'megabeds', result from catastrophic slope failures and the associated down-slope transport of enormous quantities of sediment from continental margins to the deep sea. Such large sediment failures can generate tsunamis^{2,3} and, in terrains underlain by gas hydrates (clathrates), may be associated with the release of substantial amounts of the greenhouse gas methane. It has been proposed that the megaturbidite events may be triggered by seismic

activity⁴, or may result from gas hydrate release itself^{5,6}, caused by a lowering of hydrostatic pressure on clathrates as a result of low sea level. Previous conclusions on the significance of sea-level change^{7–9}, however, have been conditional because of the lack of absolute times of turbidite emplacement. Here we use accelerator-mass-spectrometry radiocarbon dating in five widely spaced cores to constrain the date of emplacement of a large-volume (~500 km³) bed in the Balearic Basin of the western Mediterranean. This turbidite is exceptional in its magnitude and represents the main sedimentation event in the Balearic Basin over the past 100 kyr. Our data provide an estimate of 22,000 calendar years before present for emplacement of the megabed, a time when sea level stood at its lowest level during the Last Glacial Maximum. The coincidence of these dates is consistent with emplacement due to clathrate destabilization caused by low sea level, although other triggering mechanisms, such as seismic shock, cannot be ruled out.

The Balearic Abyssal Plain (Fig. 1) has an area of some 60,000 km² and is the largest plain in the Mediterranean Sea. High-resolution 3.5-kHz seismic profiles across the plain consistently show a conspicuous, thick, laterally continuous, acoustically transparent layer (Fig. 2). From such profiles, the layer is estimated to be 8–10 m thick, with its top 10–12 m below the sea floor, just too deep to be recovered fully by conventional piston coring. The layer is of consistent depth and thickness over the entire plain, except at the edges of the basin, where it onlaps the adjacent rise.

Five giant piston cores, 27–36 m in length, were taken ~100–120 km apart on the Balearic Abyssal Plain in 1995 to form an approximately north–south transect in the central plain and an east–west transect in the south of the plain (Fig. 1). The cores were all dominated by sequences of thick structureless muds (commonly grading down to basal sands and silts) separated by foraminifer-rich mud interbeds (generally <20 cm thick). The structureless muds and associated sands and silts are interpreted as turbidites on the basis of their textural characteristics, whereas the foraminifer-rich intervals are interpreted as pelagic accumulation. One particular turbidite mud, conspicuous by its thickness and with a silty to sandy base, was present in all five cores (Fig. 3). The position and thickness of this megabed is similar to that estimated for the acoustically transparent layer. Chemical analysis showed that the megabed mud has a consistent geochemical composition in all five cores, suggest-

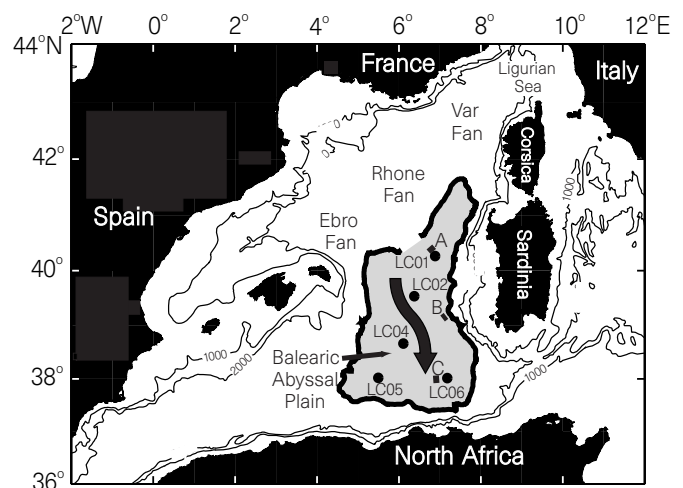


Figure 1 Map showing the positions of the five long piston cores (LC01, LC02, LC04, LC05 and LC06) recovered from the Balearic Abyssal Plain during *Marion Dufresne* cruise 81. The Balearic Abyssal Plain is defined by the 2,800-m contour, shown in bold. The distribution of the megabed is shown by a grey stipple and its emplacement direction by the black arrow. Locations of the 3.5 kHz high-resolution seismic profiles (A, B and C) illustrated in Fig. 2 are also shown. Bathymetric contours are in metres.

ing that it correlates from core-to-core. The megabed might therefore represent a single depositional event. Pelagic layers within the cores were dated biostratigraphically using the high resolution nannofossil zonal scheme of Weaver¹⁰. This shows that the sequences in the cores date back to 50–100 kyr and that the age of the megabed is less than ~50 kyr (Fig. 3). The consistent depth and thickness of the megabed suggests that it is a ponded deposit, and if present overall the entire plain it represents a volume of ~500 km³.

Determination of radiocarbon ages of pelagic beds immediately underlying and overlying the turbidite brackets its emplacement time¹¹. Such radiocarbon ages of pelagic units will be integrations over time because of bioturbation mixing by sediment fauna¹², which typically produces a surficial sediment mixed layer ~10 cm thick¹³. For this reason, a fixed cross-section of material was sampled that comprised the entire pelagic unit and the upper 10 cm of the underlying turbidite to include coarse pelagic foraminifera bioturbated down into the finer-grained turbidite since emplacement. This

yields a sample for measurement of the mean age of the entire pelagic unit, recognizing that, because of bioturbation, time resolution cannot be improved by selecting thinner samples. This procedure brackets a turbidite event most tightly when both pelagic units represent only short periods of accumulation between turbidite events.

In each of the five cores, fixed cross-section samples were taken from the first pelagic units identified immediately above and immediately below the megabed (Table 1). Clean planktonic foraminifera and pteropods were separated by hand for accelerator mass-spectrometry (AMS) radiocarbon dating from the >150 µm size fraction. This large-sized pelagic foraminifera fraction should be the least susceptible component of the sediment to current-driven redistribution. The AMS data are shown in Table 1 and Fig. 4, and our estimate for the emplacement time of the megabed is the arithmetic mean of the weighted mean ages of the overlying and underlying units, or 18,840 radiocarbon years. (Error bars represent 1 standard error).

Table 1 Radiocarbon ages for pelagic intervals bracketing Balearic Abyssal Plain megabed

Sample depth (m.b.s.f.)	Position	$\delta^{13}\text{C}$ value	$\delta^{18}\text{O}$ value	Lab. no.	^{14}C age (yr BP)	$1\sigma \pm$ error (yr BP)
LC01, 14.49–14.61	Above bed	–0.097	3.578	AA 22329	17,580	120
LC01, 23.52–23.63	Below bed	–0.152	3.738	AA 22330	20,340	150
LC02, 8.72–8.88	Above bed	0.143	3.307	AA 22331	16,900	120
LC02, 17.78–17.93	Below bed	0.53	3.528	AA 22332	23,410	210
LC04, 5.93–6.23	Above bed	–0.479	3.241	AA 22333	17,700	130
LC04, 16.96–17.08	Below bed	0.237	3.488	AA 22334	20,200	170
LC05, 5.66–5.80	Above bed	0.364	3.23	AA 22335	18,010	140
LC05, 12.11–12.27	Below bed	–2.820	–0.446	AA 22336	19,830	170
LC06, 14.46–14.65	Above bed	0.251	3.193	AA 22338	17,730	190
LC06, 20.17–20.19	Below bed	–1.254	1.735	AA 22337	34,380	660

Sample depths are given as metres below seafloor (m.b.s.f.).

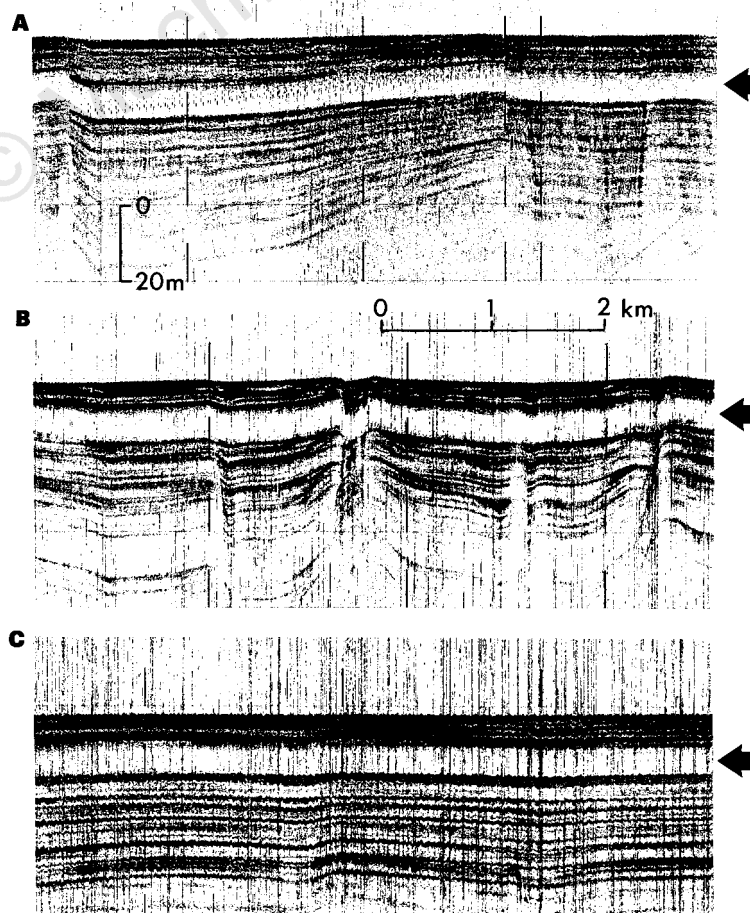


Figure 2 3.5-kHz high-resolution seismic profiles across the northern (A), central (B) and southern (C) Balearic Abyssal Plain, showing the conspicuous, thick, continuous acoustically transparent layer (arrows). Profiles are located in Fig. 1.

This estimate for the time of megabed emplacement corresponds very closely with the time of the Last Glacial Maximum and the most recent time of minimum sea level. The last glacial maximum is often quoted at ~ 18 kyr BP (ref. 14), at which time sea level stood at ~ 120 m relative to the present, which is the lowest level achieved in the past 130 kyr (ref. 15). Like our estimate, the CLIMAP¹⁴ estimate of 18 kyr was based largely on radiocarbon dating and it has subsequently been shown that conventional radiocarbon ages underestimate true time by ~ 3.0 – 3.5 kyr at 15–20 radiocarbon kyr BP (refs 16, 17). According to the radiocarbon time/calendar time relationship¹⁸, the estimate of 18,840 radiocarbon years for the time of megabed emplacement corresponds to $\sim 22,000$ calendar years. This is our best estimate of the absolute emplacement time of the megabed. It includes a 400-radiocarbon-year correction for the

age of the surface ocean mixed layer reservoir¹⁹, which also appears to be applicable to the western Mediterranean²⁰. The conclusion must be that in this case low sea level predisposed the exposed shelf and/or upper slope sediments to failure.

At the time of maximum ice sheet volume and minimum sea level around 22,000 years BP, gas hydrate (clathrate) zones worldwide would have been destabilized. As the clathrate stability field is pressure- and temperature dependent^{5,21}, the maximum clathrate instability will be at the time of lowest sea level. This destabilization could have helped trigger large-scale submarine sliding at this time^{5,21}. The physical conditions for clathrates exist in the western Mediterranean, but no bottom-simulating reflectors (BSRs) or gas hydrates have been reported²², despite extensive seismic coverage of parts of the margin. The sub-bottom depth to potential methane

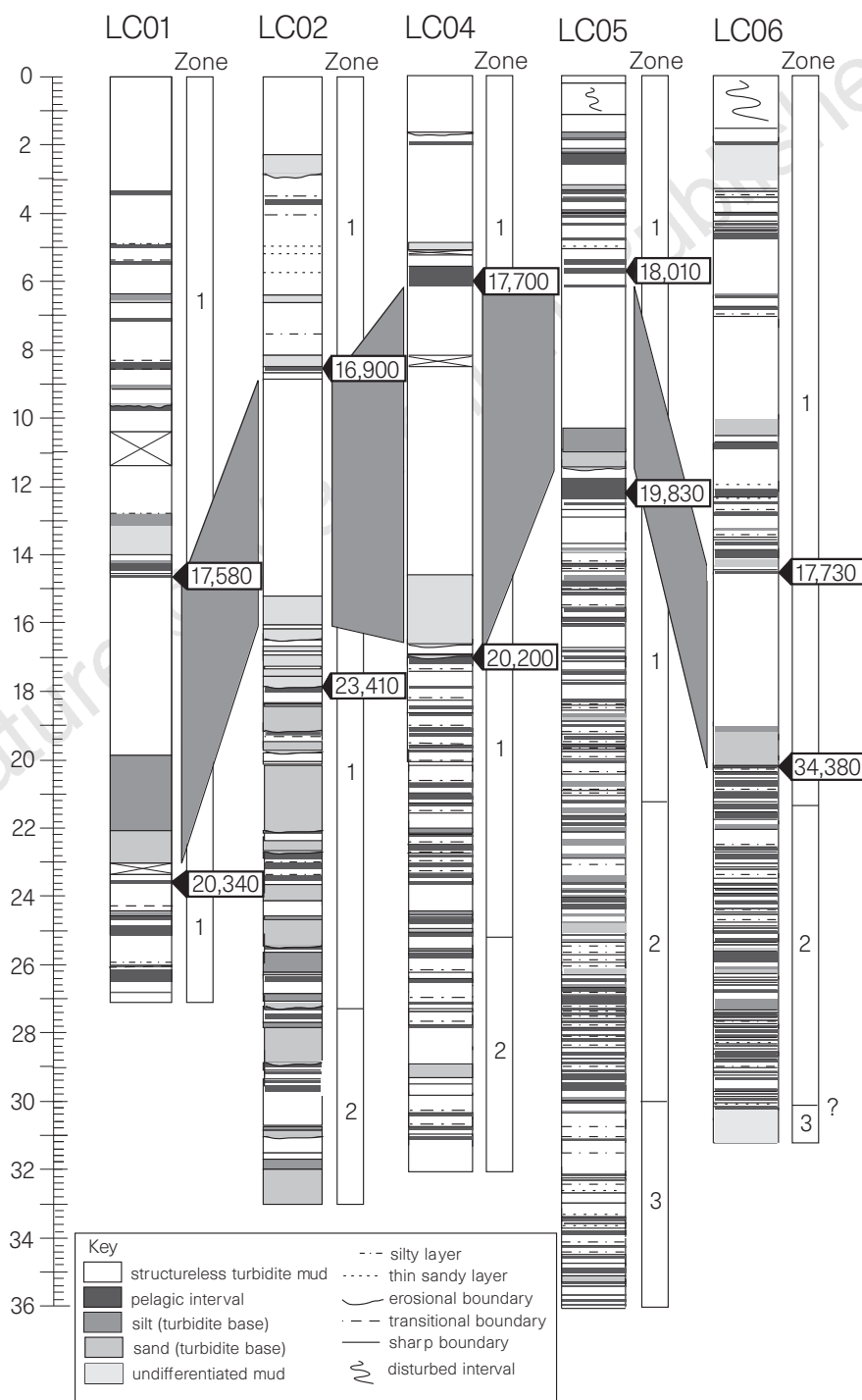


Figure 3 Graphic logs of cores LC01, LC02, LC04, LC05 and LC06 aligned side-by-side. Shown is the correlation of the thick mud megabed believed to correlate with the conspicuous transparent layer seen on 3.5-kHz seismic profiles. Note its occurrence between two distinct patterns of sedimentation: thinner, often sandier beds below, and thicker, muddier beds above. Emplacement of the megaturbidite at maximum low stand is consistent with the change in the pattern of sedimentation above and below the megabed. The thinner sandier turbidites below the megabed would have been deposited as sea level was falling when coarser-grained debris would have been deposited closer to the shelf-edge, and turbidite emplacement may have been more frequent. The thicker muddier turbidites seen above the megabed would have been emplaced as sea level was rising, when coarser-grained detritus would have been progressively trapped on the shelf. The higher position of the megabed in cores LC02, LC04 and LC05 compared with that in LC01 and LC06 probably reflects loss of the uppermost few metres in these cores during the coring process. The zones refer to the high-resolution nannofossil zonation (coccolith zones) of Weaver¹⁰, which have been tied to oxygen isotope stages. The boundary between coccolith stages 1 and 2 is dated to $\sim 50,000$ years BP. The stratigraphic position of the radiocarbon analyses reported here are also shown.

hydrate BSR on the northwestern Mediterranean margin is shallow (0–100 m)²² and previous mass wasting may have resulted in gas release. Clathrate release has been implicated in the generation of some large submarine slides, such as the Storegga Slide, off Norway, where seismic activity may have initiated the slide, and gas escape, perhaps from clathrates, may have aided sediment mobilization^{6,23}. Seismic activity on the northern Mediterranean margin is limited, but repeated seismic activity is well documented in the Ligurian Sea, north of Corsica, throughout historical time²⁴. Many of these earthquakes have epicentres on, or close to, the continental margin and appear to be linked to rejuvenation of older distensive and strike-slip structures by compressive stress related to the convergence of the European and Africa plates²⁵.

The basal sand of the megabed thickens and coarsens in grain size towards the north, suggesting emplacement from that direction. Three major fan systems are present on the northwestern Mediterranean margin—the Var, Rhone and Ebro. All drain glaciated hinterlands, and would have had high rates of sediment supply during the last glacial. The entire northwestern Mediterranean margin is prone to gravity flow processes, however, reflecting its geological setting with a mountainous borderland and rapid deposition of undercompacted sediments²⁶. The underlying Messinian salt layer may contribute to general instability by lubricating slippage of overlying sediments²⁶. The exceptional size of the Balearic Abyssal Plain megabed, however, suggests that it is an

unusual event in terms of size and frequency in the recent history of this margin and therefore may have had an extraordinary trigger. Catastrophic destabilization of margin or fan sediments, possibly due to clathrate release and/or earthquake activity, after a long period of accumulation with an increased rate of sediment supply, seems likely. □

Received 7 July 1997; accepted 27 January 1998

1. Bouma, A. H. Megaturbidite: An acceptable term? *Geo-Mar. Lett.* **7**, 63–67 (1987).
2. Dawson, A. G., Long, D. & Smith, D. E. The Storegga Slides: evidence from eastern Scotland for a possible tsunami. *Mar. Geol.* **82**, 271–276 (1988).
3. Young, R. W. & Bryant, E. A. Catastrophic wave erosion on the southeastern coast of Australia: impact of the Lanai tsunamis ca. 105 ka? *Geology* **20**, 199–202 (1992).
4. Mutti, E., Ricci Lucchi, F., Seguret, M. & Zanzucchi, G. Seismoturbidites: A new group of resedimented deposits. *Mar. Geol.* **55**, 103–116 (1984).
5. Nisbet, E. G. Sources of atmospheric CH₄ in early Postglacial time. *J. Geophys. Res.* **97**, 12859–12867 (1992).
6. Bugge, T. et al. A giant three-stage submarine slide off Norway. *Geo-Mar. Lett.* **7**, 191–198 (1987).
7. Cita, M. B. et al. Turbidites and megaturbidites from the Herodotus Abyssal Plain (eastern Mediterranean) unrelated to seismic events. *Mar. Geol.* **55**, 79–101 (1984).
8. Weaver, P. P. E. & Kuijpers, A. Climatic control of turbidite deposition on the Madeira Abyssal Plain. *Nature* **306**, 360–363 (1983).
9. Marjanac, T. Deposition of megabeds (megaturbidites) and sea-level change in a proximal part of the Eocene–Miocene flysch of central Dalmatia (Croatia). *Geology* **24**, 543–546 (1996).
10. Weaver, P. P. E. An integrated stratigraphy of the upper Quaternary of the Kings Trough flank area, NE Atlantic. *Oceanol. Acta* **6**, 451–456 (1983).
11. Thomson, J. & Weaver, P. P. E. An AMS radiocarbon method to determine the emplacement time of recent deep-sea turbidites. *Sediment. Geol.* **89**, 1–7 (1994).
12. Erlenkeuser, H. ¹⁴C age and vertical mixing of deep-sea sediments. *Earth Planet. Sci. Lett.* **47**, 319–326 (1980).
13. Cochran, J. K. Particle mixing rates in sediments of the eastern equatorial Pacific: Evidence from ²¹⁰Pb, ^{239,240}Pu and ¹³⁷Cs at MANOP sites. *Geochim. Cosmochim. Acta* **49**, 1195–1210 (1985).
14. CLIMAP Project Members. The surface of ice age earth. *Science* **191**, 1131–1137 (1976).
15. Shackleton, N. J. Oxygen isotopes, ice volume and sea level. *Quat. Sci. Rev.* **6**, 183–190 (1987).
16. Bard, E., Hamelin, B., Fairbanks, R. G. & Zindler, A. Calibration of the ¹⁴C timescale over the past 30,000 years using mass spectrometric U–Th ages from Barbados corals. *Nature* **345**, 405–410 (1990).
17. Stuiver, M. Timescales and telltale corals. *Nature* **345**, 387–388 (1990).
18. Bard, E., Arnold, M., Fairbanks, R. G. & Hamelin, B. ²³⁰Th–²³⁴U and ¹⁴C ages obtained by mass spectrometry on corals. *Radiocarbon* **35**, 191–199 (1993).
19. Bard, E. Correction of accelerator mass spectrometry ¹⁴C ages measured in planktonic foraminifera: paleoceanographic implications. *Paleoceanography* **3**, 635–645 (1988).
20. Delibrias, G. In *Nuclear Methods of Dating* (eds Roth, E. & Poty, B.) 399–436 (Kluwer Academic, Dordrecht, 1989).
21. Paull, C. K., Ussler, W. & Dillon, W. P. Is the extent of glaciation limited by marine gas-hydrates? *Geophys. Res. Lett.* **18**, 432–434 (1991).
22. Miles, P. R. Potential distribution of methane hydrate beneath the European continental margins. *Geophys. Res. Lett.* **22**, 3179–3182 (1995).
23. Evans, D., King, E. L., Kenyon, N. H., Brett, C. & Wallis, D. Evidence for long-term instability in the Storegga Slide region off western Norway. *Mar. Geol.* **130**, 281–292 (1996).
24. Giorgetti, F. & Iaccarino, E. Italian earthquake catalogue from the beginning of the Christian age up to 1968: Appendix to Giorgetti, F. & Iaccarino, E. Seismicity of the Italian Region. *Boll. Geofis. Teor. Appl.* **13**, 143–154 (1971).
25. Rehault, J.-P. & Bethoux, N. Earthquake relocation in the Ligurian Sea (western Mediterranean): Geological interpretation. *Mar. Geol.* **55**, 429–445 (1984).
26. Bellaiche, G. Sedimentary mechanisms and underlying tectonic structures of the northwestern Mediterranean margin, as revealed by comprehensive bathymetric and seismic surveys. *Mar. Geol.* **112**, 89–108 (1993).

Acknowledgements. We thank the shipboard party of *Marion Dufresne* cruise 81, in particular Y. Balut and his team for their expert core recovery, and B. Kneller for comments on the manuscript. Radiocarbon analyses were provided by the NERC Radiocarbon Laboratory, East Kilbride, and the work was partly supported by the European Union Marine Science and Technology Programme project PALAEOFLUX.

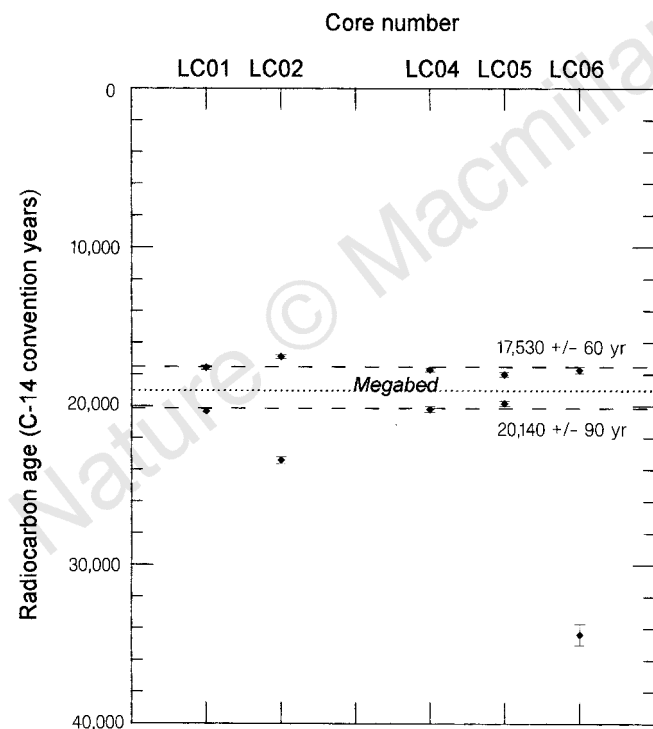


Figure 4 Radiocarbon ages derived from dating the pelagic intervals above and below the megabed in cores LC01, LC02, LC04, LC05 and LC06. The overlying pelagic samples from all five cores yield a cluster of ages between 16,900 and 18,010 radiocarbon years (Table 1), with a weighted mean age of 17,530 ± 60 radiocarbon years. There is a greater scatter in the underlying samples, with considerably older ages found for cores LC02 and LC06 than the other three. There may be unrecognized erosive effects from the megabed emplacement. The next pelagic unit down may have been selected for dating in the case of LC02. The LC06 sample was very thin (2 cm) and may have been misinterpreted as pelagic. The ages for the underlying samples in the remaining three cores are concordant, however, and yield a weighted mean age of 20,140 ± 90 radiocarbon years. The best estimate for the emplacement time of the megabed is the arithmetic mean of the weighted mean overlying and underlying ages, or 18,840 radiocarbon years.

Optimal clutch size and conspecific brood parasitism

Bruce E. Lyon

Kananaskis Field Stations, Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada, and Biology Department, Earth and Marine Sciences Building, University of California, Santa Cruz, California 95064, USA

Egg-laying organisms should lay, in a reproductive bout, the number of eggs that maximizes fitness. Lack argued 50 years ago that clutch size for most birds is limited by the amount of food parents can provide for their offspring¹. Clutch sizes, however, are often smaller than this 'Lack clutch size', and this fact is the subject of much debate^{2–4}. Here I propose and test a new explanation for this pattern that is based on evidence that conspecific brood parasitism is widespread in birds^{5–7}, specifically when females

with their own nests also parasitize conspecific birds⁸⁻¹³. A graphical model of clutch size shows that the trade-offs a brood parasite faces when allocating eggs to her own nest or to nests of other conspecific females can favour a reduction in the parasite's own clutch size. This prediction is supported by a field study of American coots (*Fulica americana*). Moreover, the cost of receiving parasitic eggs also favours a reduction in clutch size for hosts, introducing a 'game'¹⁴ element to clutch size when parasitic females are themselves parasitized. These results indicate that conspecific brood parasitism should no longer be ignored as a force in clutch-size evolution.

Nesting females who also lay eggs parasitically in the nests of conspecific birds must allocate their eggs appropriately between the nests they parasitize and their own clutch. This trade-off could explain why some females lay smaller clutches in their own nests than the anticipated Lack clutch size, as illustrated by a simple optimal clutch-size model (Fig. 1). The model can be used to examine the 'fitness increment' resulting from adding each extra egg to a female's own clutch, under the assumption that increasing clutch size yields a diminishing return on parental fitness because of constraints such as the amount of parental care needed (Fig. 1). The fitness increment of an egg is the benefit resulting from the survival of offspring minus the cost to fitness both of producing the egg and of any negative impact the egg or chick has on the survival of siblings because of competition for limited parental care.

In the absence of brood parasitism, a female reaches her optimal clutch size (N_o^*) with the addition of the last egg to her clutch that yields a positive net fitness increment (Fig. 1). The female may be physiologically capable of laying more eggs than the optimal clutch size (for example, eggs below the zero net fitness line in Fig. 1), but the fitness costs of laying the eggs and raising the young would exceed the low expected benefits^{3,4}. If this same female now has the option of laying some of her eggs parasitically in the nests of other females, how should she allocate her eggs between other nests and her own clutch? The average fitness increment she gains from laying an egg parasitically (dashed line in Fig. 1) is the 'parasitism threshold', as it determines when she should switch allocation to parasitism and, consequently, determines the new optimal clutch size for her own nest (N_p^*) when parasitism is an option: beyond this new optimal clutch size, laying eggs parasitically yields higher fitness than laying the eggs in her own nest. The term 'switch' is used loosely and the order in which parasitic and parental laying occur is

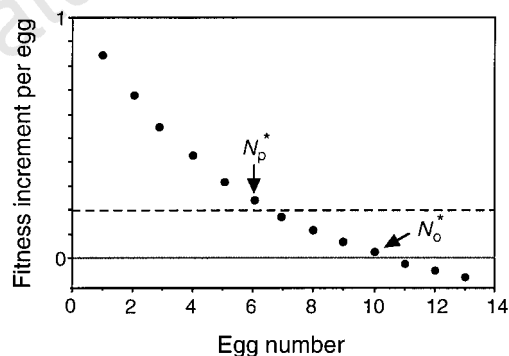


Figure 1 Model of optimal clutch size for females with, and without, the option of laying some of their eggs parasitically. The dashed line represents the average fitness per parasitic egg; eggs below this line fare better if laid parasitically rather than in the parental clutch. N_o^* is the optimal clutch size for non-parasitic females, whereas N_p^* is the optimal clutch size for brood parasites. For parasitic females, 'optimal' refers to the maximum total production of offspring in a reproductive bout (that is, parental plus parasitic young) whereas 'clutch size' is the number of eggs they lay in their own nests. Several factors can affect the average fitness of parasitic eggs relative to parental eggs^{15,21}; including when the eggs is laid in the host's breeding cycle, whether hosts reject eggs, and the number of parasitic eggs per host nest.

not important provided that females have some physiological mechanism for allocating the appropriate number of eggs to each of the two tactics.

Three main predictions for the fitness and behaviour of parasitic females derive from this model (Fig. 1). First, the fitness increment from the last egg laid to reach the optimal clutch size (egg number N_p) is greater than the average fitness gained from parasitic eggs (in rare cases, the last own egg could, by chance, yield equal fitness to parasitic eggs). Second, the average fitness gained from parasitic eggs should be greater than the fitness a female would gain from laying any additional eggs in her own nest (beyond N_p). Third, parasitism may favour a reduction in the parasite's own clutch size relative to the clutch size she would lay in the absence of parasitism (predicted reduction in Fig. 1 is $N_o^* - N_p^*$). The last prediction is particularly critical as it provides a new explanation for deviations from the Lack clutch size. A reduction in parental clutch size is predicted only when the total number of eggs a parasitic female lays is limited by her egg-laying capacity, not by her opportunities for parasitism; in other words, she has to make a trade-off.

I tested these predictions and assumptions with data from a study of parental care and brood parasitism in the American coot, a waterbird. This study took place from 1987 to 1990 near Riske Creek, British Columbia. Parasitism was very common in this population: 41% of nests were parasitized, mainly by nesting females who laid eggs parasitically immediately before starting their own nests^{12,15}. Brood parasites laid more total eggs (own plus parasitic) than did non-parasitic females¹². Moreover, the clutch sizes that parasites laid in their own nests correlated negatively with the number of eggs they laid parasitically (partial correlation, holding seasonal effects constant; $r = -0.31$, $N = 88$, $P = < 0.01$), supporting a trade-off between laying eggs parasitically and in the parental nest.

Egg survival shows a diminishing return with increasing egg position in the laying sequence (Fig. 2). This pattern results from the importance of parental feeding for chick survival, extreme hatching asynchrony and strong laying-order-dependent starvation within broods^{12,16}. Comparison of survival rates (Fig. 2) also indicates that parasites may switch between brood parasitism and laying eggs in their own nests at the appropriate clutch size, as predicted by the model; parasitic eggs (0.086 fledged chicks per egg) were less successful than last eggs laid in own nests (0.112 fledged chicks per egg), but more successful than predicted hypothetical 'next' eggs in own nests (0.063 fledged chicks per egg; survival of next eggs is predicted from a regression with the seven observed values). However, although these differences are consistent with the first two predictions of the model, they are not significant (both one-tailed $P = > 0.3$) and thus provide only qualitative support for the predictions. Given the extremely small expected (Fig. 1) and observed (Fig. 2) differences in survival among classes of marginal eggs, low statistical power makes definitive tests of the first two predictions difficult. Moreover, the observed fitness estimates do not include costs of reproduction for parental eggs or the costs associated with parasitic egg-laying, and may thus be imprecise. These two problems will probably exist in most studies. Comparisons of clutch size (the third prediction), however, should provide a more powerful and informative test of the model.

The pattern of egg survival leads to the prediction that parasitic egg-laying should be associated with a reduction in the parasite's own clutch size by one or two eggs (that is, the number of hypothetical 'next' eggs predicted to lie above the zero-fitness line but below the average fitness of parasitic eggs; Fig. 2, squares). I compared the clutch sizes of parasites with those of two different classes of non-parasitic females (Fig. 3a) that seemed to have been prevented from parasitism for different reasons. Some non-parasitic females lacked potential hosts to parasitize during their laying periods; many of these females were probably constrained by a lack of opportunity for parasitism. These females should have laid

their optimal clutch sizes (N_o^*) in their nests. In contrast, non-parasitic females who had potential hosts to parasitize were seemingly constrained by their limited fecundity. These females laid smaller clutches than the other non-parasites (Fig. 3a), supporting fecundity limitations and also suggesting that many failed to lay even their optimal clutch sizes. I therefore omitted data about these females when assessing whether brood parasites reduced their clutch sizes compared with non-parasites.

As predicted from the egg-survival data, parasites laid significantly smaller clutches than the non-parasites without hosts (Fig.

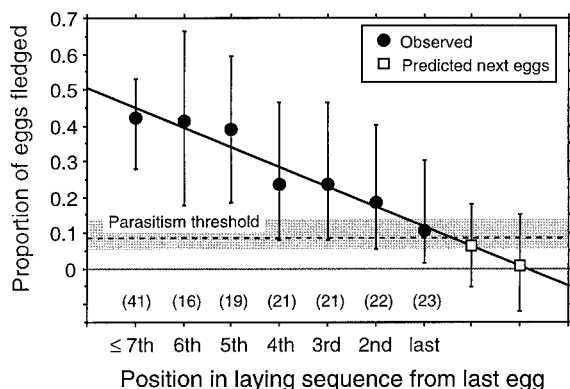


Figure 2 Observed reproductive success of eggs laid by 23 female American coots in their own nests (filled circles) as a function of position in the laying sequence, compared with the average success of 268 parasitic eggs laid by all nesting parasites (dashed line). Numbers of eggs are indicated in parenthesis. Linear regression through the seven observed data points ($F_{1,5} = 66.9$, $P = 0.0004$) predicts what females would gain by laying hypothetical 'next' eggs in the laying sequence (squares) in their own nests, rather than parasitically. Bars and shaded region denote 95% confidence intervals for parental and parasitic eggs, respectively.

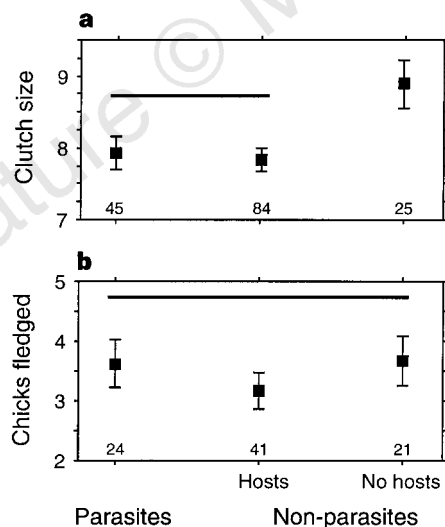


Figure 3 Comparison of clutch size and number of chicks fledged from parental nests for three classes of female American coots, namely, brood parasites, non-parasites with potential hosts to parasitize, and non-parasites that lacked potential hosts to parasitize. **a**, Mean ($\pm$ s.e.m.) clutch size; **b**, number of chicks fledged. Sample size is number of females. Horizontal bars connect means not significantly different ($P > 0.05$) according to Fisher's least significant difference posthoc tests. In theory, non-parasitic females should produce slightly more chicks than parasitic females in their own nests, because there are chicks that survive from the additional marginal eggs that non-parasites lay (Fig. 1). However, a huge sample size would be required to show whether such small differences are significant (ref. 24) (for example, a mere 2% increase in brood size is predicted on the basis of the summed survival probabilities of 'next' eggs in Fig. 2).

3a; means differ by 0.96 eggs, concurring with the predicted shift of one to two eggs). There was no difference between these two groups of females in the number of offspring fledged from their own nests (Fig. 3b), further indicating that this clutch-size difference may reflect the allocation trade-offs embodied by the model, and not simply general differences in quality between parasites and non-parasites. I believe this to be the first evidence that the competing demands of brood parasitism and own nesting for individual females can result in a shift in the optimal clutch size; these results provide a new explanation for deviations from the Lack clutch size.

The realization that brood parasitism might affect clutch size is not completely new, but the scant attention this has received previously has focused on how parasitism may affect the optimal clutch sizes of the recipients of parasitic eggs, or 'hosts'^{17–20}. My model of optimal clutch size also has implications for host clutch size. When parasitism is costly to hosts, it reduces the value of their eggs and should thus favour a reduction in their clutch size (Fig. 4a). In coots, parasitism is costly to hosts²¹, because of limited food and high starvation rates of chicks in general; hosts should therefore reduce their clutch sizes. Clutch-size comparisons confirm that hosts do adjust their clutch sizes in response to parasitism (Fig. 4b). Moreover, the amount of clutch-size reduction varied with two factors that affect the costs of parasitism²¹, namely, the number of parasitic eggs received and the timing of parasitism relative to the host's laying cycle (Fig. 4b). This is convincing evidence for adaptive clutch-size adjustment by hosts in response to parasitism^{17–20}.

An interesting consequence of adaptive clutch-size reductions for both parasites and hosts is that the optimal clutch size for parasites entails a 'game'¹⁴ element when parasitic females are themselves

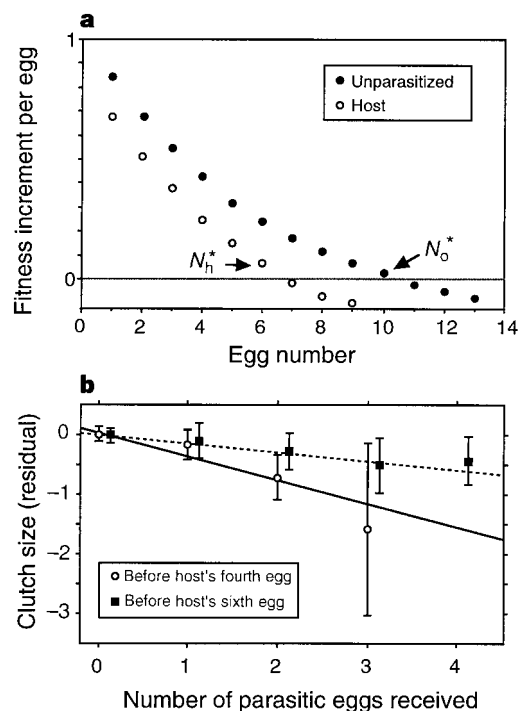


Figure 4 Adjustment of host clutch size in response to parasitism. **a**, Model of optimal clutch size for parasitized ('host') females (N_h^*) relative to unparasitized females (N_o^*). When parasitism is costly to hosts, the value of the host's eggs is reduced and a clutch-size reduction is favoured. **b**, Regressions of clutch sizes of host American coots as a function of the number of parasitic eggs they received before two different cutoff points in the laying cycle: before the host's fourth egg (solid line; $F_{1,135} = 7.36$, $P = 0.008$) and sixth egg (dashed line; $F_{1,135} = 2.15$, $P = 0.14$) were laid, respectively. Thus, host response increased with number of eggs and early parasitism: both of these factors affect the cost of parasitism and decrease the value of host eggs²¹. Regressions are based on individual nests but values shown are means and standard errors for each egg-number class.

parasitized by other females. At the population level, parasitic eggs decrease the average value of the parental eggs laid by parasites, in turn favouring even greater allocation to parasitism (trade-off in Fig. 1): optimal clutch size depends on the frequency of parasitism, and vice versa. The dynamic nature of the problem is further enhanced in species where the success of parasitic eggs is also frequency-dependent²². These various game aspects do not alter the qualitative assumptions or predictions of the graphical model. It will be important, however, to incorporate these assumptions and predictions into a more quantitative, theoretical study and into some empirical tests. For example, in populations in which parasites are also hosts, the fitness estimates for a parasite's own clutch must reflect her risks and costs of being parasitized. I have accounted for these fitness estimates; egg survival estimates (Fig. 2) included parasitic females who were themselves parasitized (6 of the 23 parasitic females). Some of these females raised parasitic chicks and, consequently, sacrificed some of their own chicks in the process.

Here I have shown that, for some species, clutch size cannot be understood without considering conspecific brood parasitism. The opposite is also true, and this clutch-size model provides a new framework for understanding brood parasitism. Most studies of parasitism do not examine clutch-size constraints and cannot explain why parasites lay eggs in the nests of others rather than in their own nests; in some cases, the hypothesis that parasitism yields a direct increase in mean fitness has been prematurely rejected²³. Earlier studies of brood parasitism now need to be reassessed. By integrating two fields of research that are generally considered in isolation of each other—study of clutch size and conspecific brood parasitism—this new clutch-size model provides a framework for enriching our understanding of both fields. □

Methods

Comparing survival of parental and parasitic eggs. Brood parasites were identified using standard techniques¹². Eggs were considered successful (fledged) if the chicks survived to 30 d after hatching¹². As it is the fitness of parental eggs relative to the fitness of parasitic eggs that is important, survival to independence is a good measure of relative fitness, assuming that post-fledging mortality is similar for both egg types. Survival rates for eggs in parasitic females' own nests were calculated for successful nests (some eggs hatched) but then adjusted by the proportion of parental eggs that were laid in successful nests (82.1% of 731 eggs). Only nests where the fates of more than half the chicks were known were included in these analyses. Confidence limits for proportion of eggs surviving (Fig. 2) were based on sample size²⁴ for observed eggs and on 1,000 bootstrapped regression equations for predicted next eggs. For statistical comparisons of egg survival, a G-test compared last parental eggs with parasitic eggs, whereas 1,000 bootstrapped predictions of 'next'-egg survival were used to compare 'next' eggs with parasitic eggs.

I examined egg success in relation to an egg's position in the laying sequence (backwards from the last egg laid in the clutch), rather than on the basis of clutch size, to enable pooling of results despite large variations in clutch size¹² and to predict what parasites would gain if they were to add 'hypothetical' next eggs in the laying sequence to their clutches (Fig. 2). The survival value of eggs from a specific position in the laying sequence would not indicate the fitness increments from those eggs if later-laid eggs survive at the expense of earlier-laid eggs, because the survival of the eggs would need to be discounted by the fitness reduction they caused through the death of siblings. However, the two measures (survival and fitness increment) will be equivalent where there is strict laying-order-dependent survival within broods; in this study, few later-laid eggs survived at the expense of earlier-laid ones^{12,16}.

Clutch- and brood-size comparisons. For analysis of sizes of clutches of parasitic and non-parasitic females, host availability was determined on the basis of the observed spatial and temporal patterns of host use¹⁵. To reduce variance due to strong seasonal decline in clutch size, the effects of date were controlled by analysis of covariance (ANCOVA) ($F = 9.04$, $P = 0.002$); clutch sizes (Fig. 3) are therefore adjusted means. For chicks, the assumptions of ANCOVA are violated, so analysis of variance was used ($F = 0.64$, $P = 0.53$)

but, to avoid bias due to differences among groups in timing of nesting, only birds who initiated laying within 20 d of the first egg laid in the population are included.

To avoid potential confounding effects of female quality in the analysis of host clutch-size responses, only host nests were included in these analyses. Residual clutch sizes from regressions of clutch size against laying data were used to control for strong seasonal declines in clutch size²¹.

Received 30 May; accepted 11 December 1997.

1. Lack, D. The significance of clutch size. *Ibis* **89**, 309–352 (1947); *Ibis* **90**, 25–45 (1948).
2. Klomp, H. The determination of clutch size in birds, a review. *Ardea* **58**, 1–124 (1970).
3. Martin, T. E. Food as a limit on breeding birds: a life history perspective. *Annu. Rev. Ecol. Syst.* **18**, 453–487 (1987).
4. Godfray, H. C. J., Partridge, L. & Harvey, P. H. Clutch size. *Annu. Rev. Ecol. Syst.* **22**, 409–429 (1991).
5. Yom-Tov, Y. Intraspecific nest parasitism in birds. *Biol. Rev.* **55**, 93–108 (1980).
6. Andersson, M. In *Producers and Scroungers* (ed. Barnard, C. J.) 195–228 (Croon Helm, London, 1984).
7. Rohwer, F. C. & Freeman, S. The distribution of conspecific nest parasitism in birds. *Can. J. Zool.* **67**, 239–253 (1989).
8. Brown, C. R. Laying eggs in a neighbor's nest: benefit and cost of colonial nesting in swallows. *Science* **224**, 518–519 (1984).
9. Gibbons, D. W. Brood parasitism and cooperative nesting in the moorhen, *Gallinula chloropus*. *Behav. Ecol. Sociobiol.* **19**, 221–232 (1986).
10. Møller, A. P. Intraspecific nest parasitism and anti-parasite behaviour in swallows *Hirundo rustica*. *Anim. Behav.* **35**, 247–254 (1987).
11. Emlen, S. T. & Wrege, P. H. Forced copulation and intra-specific parasitism: two costs of social living in the white-fronted bee-eater. *Ethology* **71**, 2–29 (1986).
12. Lyon, B. E. Brood parasitism as a flexible female reproductive tactic in American cots. *Anim. Behav.* **46**, 911–928 (1993).
13. Jackson, W. M. Causes of conspecific nest parasitism in the northern masked weaver. *Behav. Ecol. Sociobiol.* **32**, 119–126 (1993).
14. Maynard Smith, J. M. *Evolution and the Theory of Games* (Cambridge Univ. Press, 1982).
15. Lyon, B. E. Tactics of parasitic American cots: host choice and the pattern of egg dispersion among host nests. *Behav. Ecol. Sociobiol.* **33**, 87–100 (1993).
16. Lyon, B. E., Eadie, J. M. & Hamilton, L. D. Parental preference selects for ornamental plumage in American cot chicks. *Nature* **371**, 240–243 (1994).
17. Andersson, M. & Eriksson, M. O. G. Nest parasitism in goldeneyes *Bucephala clangula*: some evolutionary aspects. *Am. Nat.* **120**, 1–16 (1982).
18. Power, H. W. et al. The parasitism insurance hypothesis: why starlings leave space for parasitic eggs. *Condor* **91**, 753–765 (1989).
19. Eadie, J. M. *Alternative Reproductive Tactics in a Precocial Bird: the Ecology and Evolution of Brood Parasitism in Goldeneyes*. Thesis, Univ. British Columbia (1989).
20. Rothstein, S. I. Brood parasitism and clutch-size determination in birds. *Trends Ecol. Evol.* **5**, 101–102 (1990).
21. Lyon, B. E. *The Ecology and Evolution of Conspecific Brood Parasitism in American Cots* (Fulica americana). Thesis, Princeton Univ. (1992).
22. Eadie, J. M. & Fryxell, J. M. Density-dependence, frequency-dependence, and alternative nesting strategies in goldeneyes. *Am. Nat.* **140**, 621–641 (1992).
23. Brown, C. & Brown, M. B. A new form of reproductive parasitism in cliff swallows. *Nature* **331**, 66–68 (1988).
24. Sokal, R. R. & Rohlf, F. J. *Biometry* (W. H. Freeman, San Francisco, 1981).

Acknowledgements. I thank P. Grant, D. Rubenstein, C. Martinez del Rio and H. Horn for advice during the study; R. Cartar, J. Eadie, D. Hill, A. Kacelnik and T. Martin for comments on the manuscript; and L. Hamilton, B. Bair, L. Cargill, S. Everding, D. Hansen, M. Magrath and C. Morrill for help in the field. Research was funded by the Chapman Fund, National Geographic Society, National Science Foundation, Princeton University and Sigma Xi Society. Kananaskis Field Stations of the University of Calgary provided support during writing.

Correspondence and requests for materials should be addressed to B.E.L. at Santa Cruz (e-mail: lyon@biology.ucsc.edu).

Exceptional soft-tissue preservation in a theropod dinosaur from Italy

Cristiano Dal Sasso* & Marco Signore†‡

* Museo Civico di Storia Naturale, Corso Venezia 55, 20121 Milano, Italy

† Dipartimento di Paleontologia, Università degli Studi di Napoli "Federico II", Largo S. Marcellino 10, 80138 Napoli, Italy

‡ Department of Geology, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK

The Lower Cretaceous Pietraroia Plattenkalk (Benevento Province, southern Italy) has been known since the eighteenth century for its beautifully preserved fossil fishes. During Albian time (about 113 Myr ago¹), deposition of fine marly limestone in a shallow lagoonal environment, affected by cyclic periods of low oxygen levels², led to exceptional preservation of soft tissue in a



Figure 1 The holotype of *Scipionyx samniticus*, gen. et sp. nov., fossilized in a beige limestone from the Lower Cretaceous (Albian) of Pietraroia (Benevento, southern Italy). Scale bar, 2 cm. Courtesy of the Soprintendenza Archeologica, Salerno.

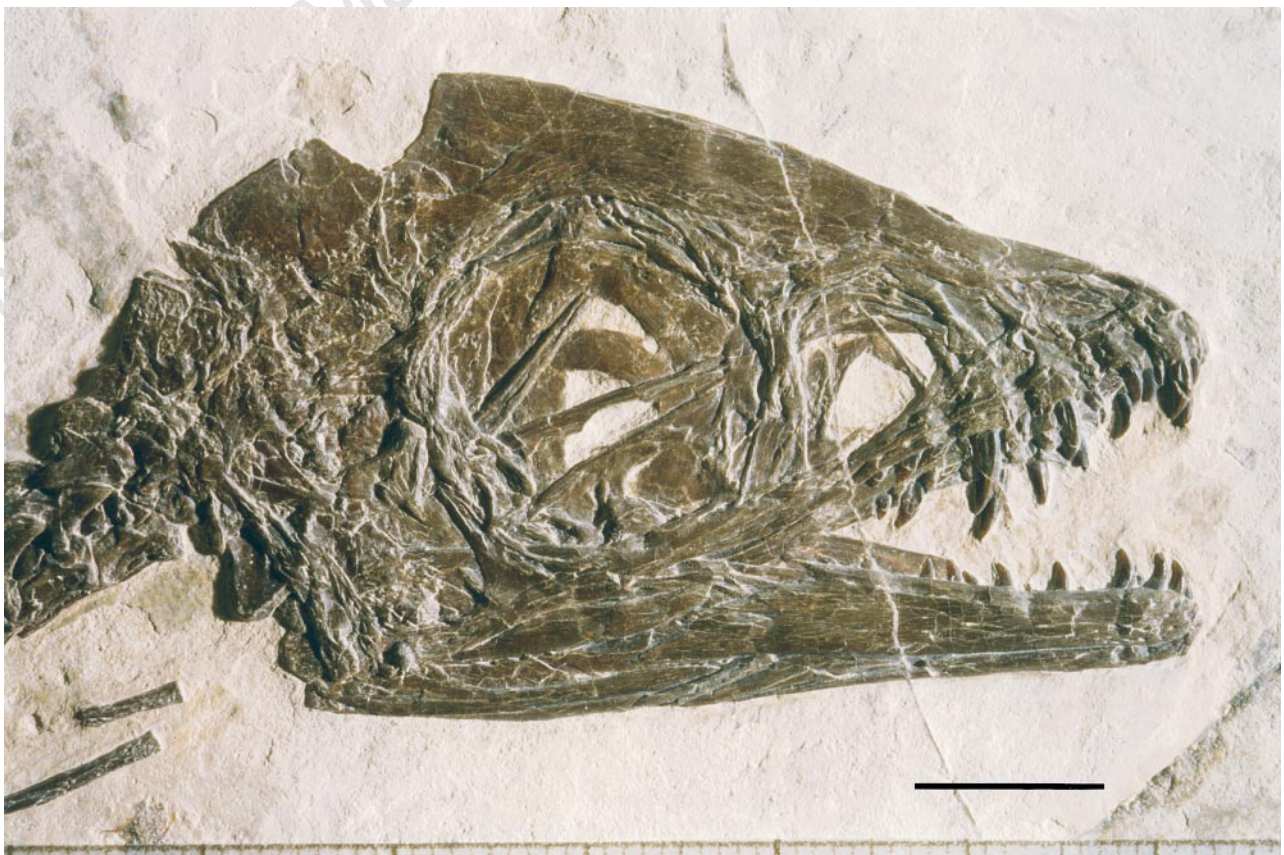
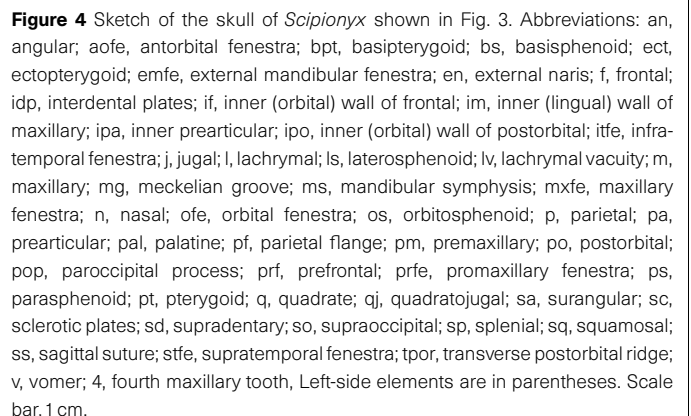
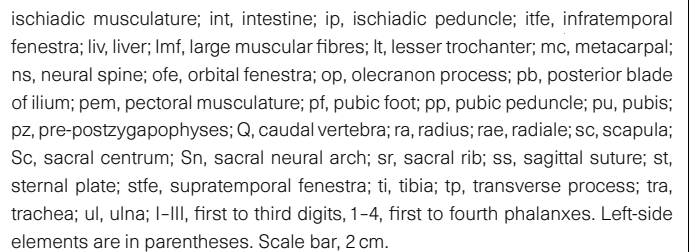


Figure 3 Close-up of the skull of *Scipionyx*. Scale bar, 1 cm. Courtesy of the Soprintendenza Archeologica, Salerno.

The skeleton of the Pietraröia dinosaur, 237-mm long from the tip of premaxilla to the last (ninth) preserved caudal vertebra, lies on its left side (Figs 1, 2), in nearly perfect anatomical articulation. Although the head is upturned with respect to the position in life,



there is no opisthotonic condition⁹ in the neck. The hindlimbs are missing distal to the proximal epipodials, as are most of the tail and most of the second right manual claw.

Theropoda

Tetanurae

Coelurosauria

Maniraptoriformes

Scipionyx samniticus gen. et sp. nov.

Etymology. Scipio (Latin male name): dedicated to Scipione Breislak, who first described the Pietraroia Plattenkalk, and Publius Cornelius Scipio (nicknamed Africanus), *consul militaris* of the Roman Army, who fought in the Mediterranean area; onyx (Greek): claw; samniticus (Latin): of the Samnium, ancient name of the region including the Benevento Province.

Holotype. Nearly complete, articulated skeleton, housed at the Soprintendenza Archeologica, Salerno.

Horizon and locality. Lower Cretaceous (Albian) of the Pietraroia Plattenkalk (Benevento Province, southern Italy).

Diagnosis. Referable to Theropoda^{10–12} for the synapomorphic presence of denticulate teeth, intramandibular joint, straplike scapular blade, distal carpal 1 clasping metacarpals I and II, manus with digits IV and V absent and with elongate penultimate phalanges, booted pubis. Referable to Coelurosauria^{11,12} on the basis of the

derived presence of jugal participation in the antorbital fenestra and metacarpal I being one-third the length of metacarpal II; more primitive than other coelurosaurians in retaining stout lachrymal and ischial foot. Referable to Maniraptoriformes^{12,13} on the basis of the derived presence of third antorbital fenestra, elongate cervical prezygapophyses, forelimb/presacral ratio of 0.75, ulna bowed posteriorly, semilunate carpal, and slender metacarpal III. Differs from all other Maniraptoriformes in the unique possession of an accessory transverse postorbital ridge at fronto-parietal contact, and by the compressed nature of the radiale and semilunate carpal; differs also in the primitive retention of large prefrontal, pronounced scapular acromion, and rounded coracoid caudal end.

The body proportions indicate that this animal is little more than a hatchling^{14–16}. The skull/presacral ratio (0.48) is higher than in any known adult theropod, including *Tyrannosaurus*¹⁷; moreover, the antorbital region is short, and the orbit is large and circular. Many skeletal elements (scapulo-coracoid, sternal plates, sacral vertebrae) are unfused¹⁸, and several neural arches are separate from their centra. The symmetry of tooth development in both maxillary rami suggests that the first tooth replacement had not occurred; furthermore, the low denticle count may be related to juvenile age^{14,19}.

A combination of characters that usually identify different clades^{11,12} is present in the Italian theropod. The forelimb ratios and most elements of the skull (Figs 3, 4) resemble features of dromaeosaurids^{17,20,21}. Among them are the following derived characters: sloping postorbital region; quadrate with a single head

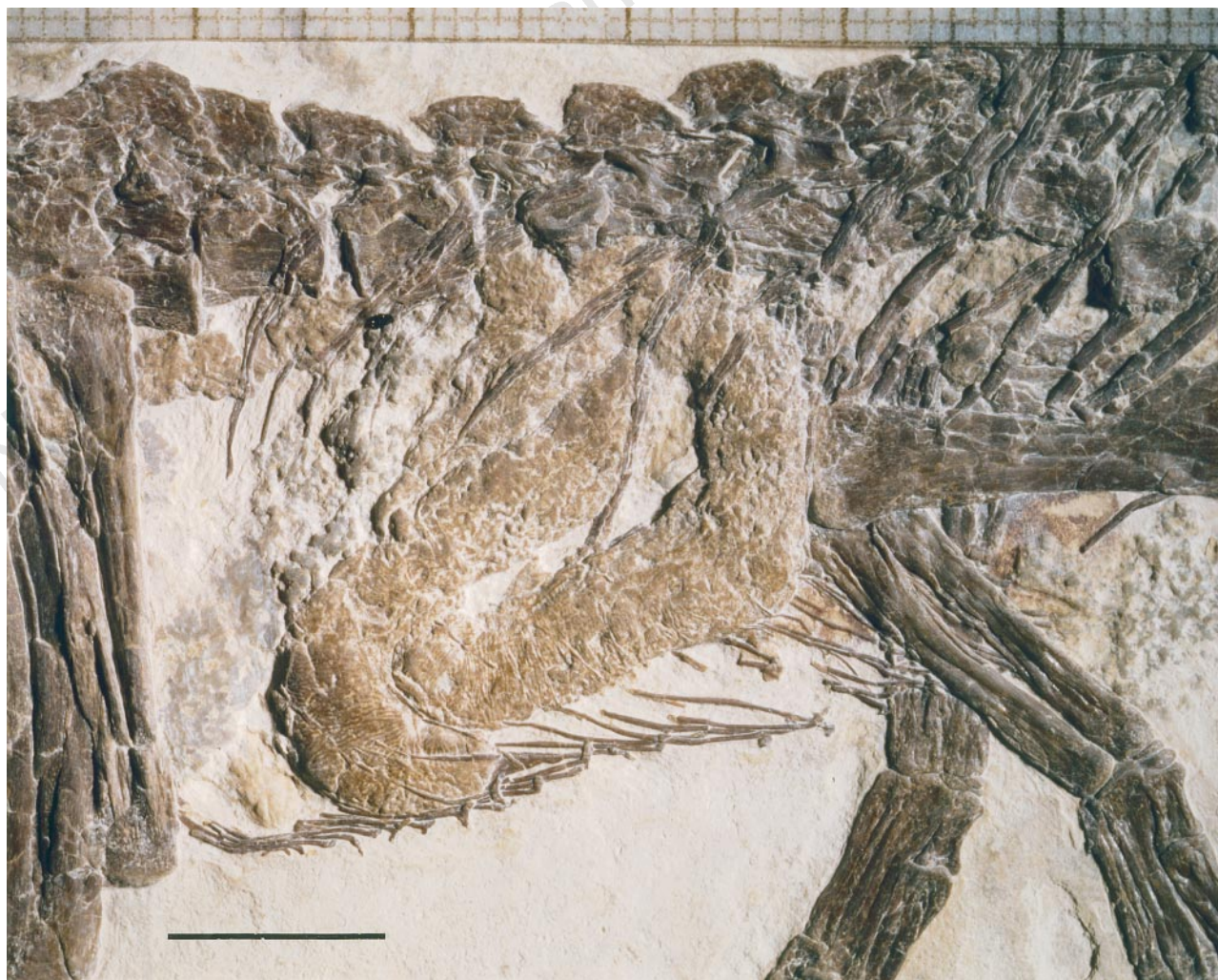


Figure 5 Close-up of the abdomen of *Scipionyx*, showing the perfectly fossilized intestine (left) and a reddish macula that might represent the remains of the liver

(right). Scale bar, 1 cm. Courtesy of the Soprintendenza Archeologica, Salerno.

articulating exclusively with the squamosal; paroccipital processes distally slightly twisted postero–dorsally; maxillary fenestra dorsally displaced; palatine pneumatization; alignment of dentary foramina into two rows, not in grooves; and splenial emerging externally on the lateral surface of the mandible. Other characters, such as the small, L-shaped quadratojugal with equal rami, low cervical neural spines, and the barely propubic pelvis, are not dromaeosaurid, but appear synapomorphic with the Troodontidae²². There are also ornithomimid-like, plesiomorphic features, clearly differentiating the pelvis of *Scipionyx* from both the dromaeosaurid and the troodontid pattern (for example, the ilium is posteriorly truncated, and the ischium is three-quarters of the pubic length, with a forward-pointing foot^{12,23}). Moreover, characters such as the L-shaped lachrymal, fan-like coracoid, feeble development of the fourth trochanter, and slender and slightly curved chevrons, resemble a more generalized coelurosaurian bauplan^{9,11,24}.

The mosaic of characters does not allow attribution of this new genus to any known theropod family. But the phylogenetic relationships of *Scipionyx* must undoubtedly be searched for within the Maniraptoriformes, as it shares at least six unambiguous synapomorphies with that clade^{12,13} (see Diagnosis).

A unique, striking feature of the specimen is the preservation of soft parts (Figs 1, 5). Muscles are present in the pectoral area, with scattered acicular fibres clearly visible under $\times 50$ magnification. At the tail base, a fascium with at least three different arrangements of fibrae longae possibly represents part of the M. caudifemoralis longus^{25,26}. Most of the intestine (tenuis²⁷), 5.22-mm average diameter) is positioned further forwards than it is generally thought to be^{25,28}, whereas the colon²⁷ passes through the pelvic canal, close to the vertebral column, and ends just above the ischiadic foot. The gut is surprisingly short and deep in section, suggesting a high absorption rate. Its muscular wall has transverse folds, which are sometimes anastomized. Immediately above the furcula, there appear to be some tracheal rings. A large, reddish, well delimited haematitic halo is tentatively interpreted as liver traces, mainly because of its post-sternal location.

The gastralia, still in life position, allow estimation of the abdominal depth and reveal their contribution to an effective support for the posterior intestinal tract. The presence of a furcula in this articulated specimen eliminates every doubt about the interpretation of similar structures in other theropods^{29,30}. □

Received 17 September 1997; accepted 9 February 1998.

- Bravi, S. & De Castro, P. The Cretaceous fossil fishes level of Capo d'Orlando, near Castellammare di Stabia (NA): biostratigraphy and depositional environment. *Mem. Sci. Geol.* **47**, 45–72 (1995).
- Bravi, S. *Contributo allo studio del giacimento ad ittioliti di Pietraroja (Benevento)* (Thesis, Dip. Geol. Univ. Napoli "Federico II", 1987).
- Leonardi, G. & Teruzzi, G. Prima segnalazione di uno scheletro fossile di dinosauro (Theropoda, Coelurosauria) in Italia (Cretacico di Pietraroja, Benevento). *Paleoconache* **1993**, 7–14 (1993).
- Leonardi, G. & Avanzini, M. Dinosauri in Italia. *Le Scienze (Quaderni)* **76**, 69–81 (1994).
- Signore, M. *Il teropode del Plattenkalk della Civita di Pietraroja (Cretaceo inferiore, Bn)* (Thesis, Dip. Paleont. Univ. Napoli "Federico II", 1995).
- Seilacher, A. Begriff und Bedeutung der Fossil-Lagerstätten. *N. Jb. Geol. Paläont. Mh. Jahr.* **1970**, 34–39 (1970).
- Kellner, A. Fossilized theropod soft tissue. *Nature* **379**, 32 (1996).
- Ji, Q. & Ji, S. On discovery of the earliest bird fossil in China and the origin of birds. *Chinese Geol.* **233**, 30–33 (1996).
- Ostrom, J. H. The osteology of *Compsognathus longipes* Wagner. *Zitteliana* **4**, 73–118 (1978).
- Benton, M. J. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmólska, H.) 11–30 (Univ. California, Berkeley, 1990).
- Gauthier, J. in *Mem. Calif. Acad. Sci.* (ed. Padian, K.) **8**, 1–55 (1986).
- Holtz, T. R. Jr The phylogenetic position of the Tyrannosauridae: implications for theropod systematics. *J. Paleontol.* **68**, 1100–1117 (1994).
- Holtz, T. R. Jr Phylogenetic taxonomy of the Coelurosauria (Dinosauria: Theropoda). *J. Paleontol.* **70**, 536–538 (1996).
- Norell, M. A. et al. A theropod dinosaur embryo and the affinities of the Flaming Cliffs dinosaur eggs. *Science* **266**, 779–782 (1994).
- Colbert, E. H. in *Dinosaur Systematics* (eds Carpenter, K. & Currie, P. J.) 81–90 (Cambridge Univ. Press, 1990).
- Molnar, R. E. in *Dinosaur Systematics* (eds Carpenter, K. & Currie, P. J.) 71–79 (Cambridge Univ. Press, 1990).
- Ostrom, J. H. Osteology of *Deinonychus antirrhopus*, an unusual theropod from the Lower Cretaceous of Montana. *Bull. Peabody Mus. Nat. Hist.* **30**, 1–165 (1969).
- Currie, P. J. & Zhao, X. A new carnosaur (Dinosauria, Theropoda) from the Jurassic of Xinjiang, People's Republic of China. *Can. J. Earth Sci.* **30**, 2037–2081 (1993).

- Currie, P. J. et al. in *Dinosaur Systematics* (eds Carpenter, K. & Currie, P. J.) 107–125 (Cambridge Univ. Press, 1990).
- Currie, P. J. New information on the anatomy and relationships of *Dromaeosaurus albertensis* (Dinosauria: Theropoda). *J. Vert. Paleontol.* **15**, 576–591 (1995).
- Witmer, L. M. & Maxwell, W. D. The skull of *Deinonychus* (Dinosauria: Theropoda): new insights and implications. *J. Vert. Paleontol.* **16**, 73A (1996).
- Russell, D. A. & Dong, Z. A nearly complete skeleton of a new troodontid dinosaur from the Early Cretaceous of the Ordos Basin, Inner Mongolia, People's Republic of China. *Can. J. Earth Sci.* **30**, 2163–2173 (1993).
- Barsbold, R. & Osmólska, H. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmólska, H.) 225–244 (Univ. California, Berkeley, 1990).
- Norman, D. B. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmólska, H.) 280–305 (Univ. California, Berkeley, 1990).
- Paul, G. S. *Predatory Dinosaurs of the World* 1–464 (Simon & Schuster, 1988).
- Gatesy, S. M. in *Functional Morphology in Vertebrate Paleontology* (ed. Thomason, J.) 219–234 (Cambridge Univ. Press, 1995).
- Guibé, J. *Traité de Zoologie* 14–II (ed. Grassé, P. P.) 521–548 (Masson, 1970).
- Wellnhofer, P. in *The Beginning of Birds* (eds Hecht, M. K., Ostrom, J. H., Viohl, G. & Wellnhofer, P.) 113–122 (Freunde des Jura-Museums, Eichstätt, 1985).
- Bryant, H. N. & Russell, A. P. The occurrence of clavicles within Dinosauria: implications for the homology of the avian furcula and the utility of negative evidence. *J. Vert. Paleontol.* **13**, 171–184 (1993).
- Norell, M. A. et al. A *Velociraptor* wishbone. *Nature* **389**, 447 (1997).

Acknowledgements. We thank P. J. Currie and M. J. Benton for critical revision of the manuscript; C. Barbera, L. Chiappe, P. Makovicky, M. Norell, M. Novacek, E. Koppelhus, E. Nicholls, A. Kotsakis, D. Maxwell, R. Molnar, G. Pasini, E. Signore, G. Todesco, and L. Witmer; G. Leonardi for his encouragement; S. Rampinelli for the superb preparation of the specimen; G. Teruzzi, Dinosaur Project coordinator; and G. Tocco for allowing us to study the specimen. Photographs are by L. Vitola; drawings are by F. Fogliazza.

Correspondence and requests for materials should be addressed to C.D.S. (e-mail: cdalasso@yahoo.com).

Somatosensory discrimination based on cortical microstimulation

Ranulfo Romo, Adrián Hernández, Antonio Zainos & Emilio Salinas

Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, 04510 México DF, México

The sensation of flutter is produced when mechanical vibrations in the range of 5–50 Hz are applied to the skin^{1–3}. A flutter stimulus activates neurons in the primary somatosensory cortex (S1) that somatotopically map to the site of stimulation^{4,5}. A subset of these neurons—those with quickly adapting properties, associated with Meissner's corpuscles—are strongly entrained by periodic flutter vibrations, firing with a probability that oscillates at the input frequency^{1,6}. Hence, quickly adapting neurons provide a dynamic representation of such flutter stimuli. However, are these neurons directly involved in the perception of flutter? Here we investigate this in monkeys trained to discriminate the difference in frequency between two flutter stimuli delivered sequentially on the fingertips^{1,7}. Microelectrodes were inserted into area 3b of S1 and the second stimulus was substituted with a train of injected current pulses. Animals reliably indicated whether the frequency of the second (electrical) signal was higher or lower than that of the first (mechanical) signal, even though both frequencies changed from trial to trial. Almost identical results were obtained with periodic and aperiodic stimuli of equal average frequencies. Thus, the quickly adapting neurons in area 3b activate the circuit leading to the perception of flutter. Furthermore, as far as can be psychophysically quantified during discrimination, the neural code underlying the sensation of flutter can be finely manipulated, to the extent that the behavioural responses produced by natural and artificial stimuli are indistinguishable.

Two monkeys (*Macaca mulatta*) were trained in a standard discrimination task⁷ in which two mechanical vibrations, termed base and comparison, are delivered in each trial (Fig. 1). The monkeys learned to indicate whether the comparison stimulus

was of higher or lower frequency than the base. After training, neurophysiological experiments were done while the task was performed. In every session, single and multiunit neuronal activity was recorded through microelectrodes inserted into area 3b of S1 (ref. 1). We needed small cutaneous receptive fields that were confined to the smooth, hairless skin of one fingertip and possessed quickly adapting (QA) properties¹; the stimulator tip was placed on the centre of the receptive fields. We then switched to the electrical microstimulation protocol, in which the stimulation frequencies and event sequence remained the same, but which substituted the second mechanical stimulus with a train of current bursts injected into area 3b (Fig. 1). The bursts were delivered at the comparison frequency. As S1 is organized into modules of neurons with similar functional properties (columns)^{5,8,9}, we aimed to drive a cluster of 3b neurons—mostly of the QA type, hopefully—in a way that matched the periodic responses seen with mechanical stimuli, and to find out whether the animals could interpret the artificial signals as flutter.

It took several days to find correct microstimulation parameters for the first monkey, but thereafter the parameters were rarely modified and both animals were consistently correct 75% of the time. The monkeys reacted quickly enough and were able to discriminate between base and comparison optimally when each burst consisted of two current pulses of amplitudes of $>65 \mu\text{A}$. A typical stimulus set included 10 or 20 frequency pairs (of base and comparison). In each data collection run, a frequency pair was chosen randomly in every trial until 10 trials per pair had accumulated. Microstimulation trials were randomly intermixed with standard, purely mechanical discrimination trials. The base and comparison frequencies varied across the same range in these two types of trial. Figure 2a shows the results obtained using a set of 10 frequency pairs with a constant difference of 8 Hz. Results shown in Fig. 2b, c were obtained with another stimulus set for which traditional psychometric curves could be fitted to the points of equal base. The results are almost identical for standard and microstimulation trials from the same runs. Including all 30 frequency pairs in Fig. 2, discrimination with mechanical and electrical comparisons was, on average, 85% and 83% correct, respectively; the difference was not significant (permutation test¹⁰, $n = 240$, $P > 0.19$). The monkeys were consistently able to extract the comparison frequency from the artificially induced sensation.

Interestingly, with current amplitudes of $<40 \mu\text{A}$, the monkeys just waited, as if no comparison stimulus had been applied. When currents of between 40 and $65 \mu\text{A}$ and/or single-pulse bursts were used, the monkeys could detect the artificial stimulus—their

reaction times were normal—but performance levels were near chance levels (in 18 runs). They behaved as if the stimuli had fallen inside the so-called atonal interval, a range of amplitudes in which flutter vibrations can be detected but their frequencies cannot be resolved¹¹. In seven runs, a high-frequency microstimulation current (of 50–200 Hz) was superimposed on the mechanical comparison stimulus, almost always overriding it: in 348 of 400 such trials, the comparison was lower than the base but was called higher. In 69 runs, the animals successfully discriminated between artificial and natural stimuli. However, in 12 runs they could not discriminate, even though effective microstimulation parameters were used. This occurred when the stimulating electrode was placed around neurons with slowly adapting properties, or when it penetrated the white matter below area 3b; however, in this case accurate discriminations were made when the electrode was again positioned near QA neurons in area 3b.

It has been proposed that discrimination of flutter stimuli depends on the periodic structure of the spike trains evoked in S1 (refs 1, 3). Alternatively, this structure might simply reflect the periodicity of the input, without being essential for the discrimination process itself. This would be the case if, for example, discrimination were based on the mean number of spikes (or bursts¹²) fired by the population of QA neurons as a function of stimulus frequency. We investigated these possibilities in experiments in which the comparison stimulus was aperiodic, but of the same mean frequency as the periodic stimulus. In mechanical-stimulation experiments the comparison was periodic in one-half of the

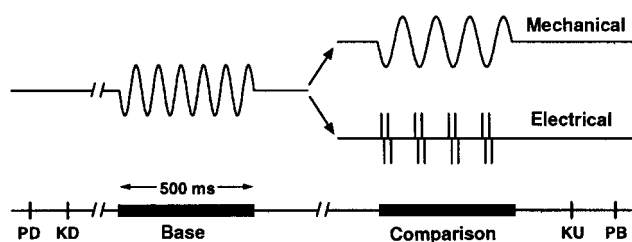


Figure 1 Sequence of events during standard (up arrow) and microstimulation (down arrow) discrimination trials. The mechanical probe is lowered, indenting the glabrous skin of one digit of the restrained hand (PD); the monkey places his free hand on an immovable key (KD) within 1 s after indentation; after a delay period (1.5–3 s) the probe oscillates vertically, sinusoidally, at the base frequency; after an interstimulus interval (1–3 s) either a second mechanical stimulus or a train of microstimulation bursts is delivered at the comparison frequency; the monkey releases the key (KU) within 600 ms and presses one of two push-buttons (PB) to indicate whether the comparison frequency was higher or lower than the base.

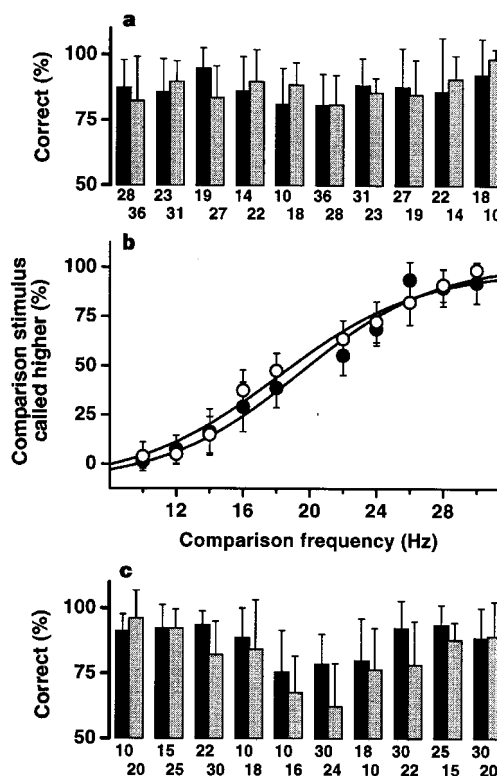


Figure 2 Psychophysical performance using periodic stimuli. Black bars and circles indicate standard discrimination trials; grey bars and open circles indicate microstimulation trials. Upper and lower rows of numbers on the x axis in **a**, **c** indicate, respectively, base and comparison frequencies for each pair of bars in Hz. **a**, Results from one stimulus set with ten frequency pairs. **b**, Data and sigmoidal fits (χ^2 test, $P < 0.001$) for ten pairs in which the base frequency was 20 Hz. **c**, Results from ten other pairs randomly intermixed with those in **b**. Each data point represents 80 trials, 50 from monkey 1 and 30 from monkey 2, collected in 8 uninterrupted runs. Error bars are 1 s.d. with respect to the 8 runs.

trials—the standard situation—and in the other half it consisted of n single-cycle sinusoids (n = stimulation period $\times$ comparison frequency) of a small, fixed width (20 ms), separated by random time intervals. In microstimulation experiments, again half of the trials were standard—the comparison was mechanical and periodic—but in the other half of the trials microstimulation current was injected; it consisted of n bursts with random time intervals between them. In all cases the base was mechanical and periodic. Figure 3 shows the results obtained with the two types of aperiodic comparison stimuli, using the same frequency sets as in Fig. 2. No previous training had taken place with these novel stimuli. Two facts are notable: both animals were immediately able to extract the mean frequency from the non-periodic signals, and they did this reliably and almost identically in the two situations. For the 30 frequency pairs shown in Fig. 3, discrimination with mechanical and electrical comparisons was, on average, 84% and 81% correct, respectively; the small difference was significant (permutation test¹⁰, n = 240, P < 0.001). Human subjects could easily distinguish between the periodic and aperiodic conditions, indicating that these conditions correspond to slightly different tasks (Romo *et al.*, unpublished observations). Nevertheless, the results suggest that the exact values of the interspike (or interburst) intervals are not essential for mean-frequency discrimination. The detailed structure of the interspike intervals may still be important for other tasks, such as recognition of mechanical patterns.

In an extra control experiment we investigated the effect of stimulus amplitude, which could potentially act as an alternate

discrimination signal⁷. Four frequency pairs were used, and all stimuli were periodic. In one-half of the trials the two stimuli were mechanical, and in the other half the comparison stimulus was electrical. In both cases, the amplitude of the comparison could take one of three values, a standard amplitude A , $0.6A$, or $1.4A$. The changes in intensity were, in terms of percentage, of the same magnitude (40%) as the differences between base and comparison frequencies. Figure 4 shows that performance was not affected by the large variations in amplitude (permutation tests¹⁰, n = 12, P > 0.03); had the monkeys been guided by the amplitude changes, one of the three combinations for each frequency pair would have fallen to $\leq 25\%$ correct, because performance was normally >75% correct. Amplitude corrections, like those of mechanical stimuli (see Methods), were also applied to electrical currents in $\sim 60\%$ of all runs; they had no marked impact on performance.

The precise effect of microstimulation is hard to assess. Is it possible that monkeys learned to discriminate a feature in the artificial signal other than frequency? This is unlikely. First, flutter vibrations have essentially two parameters, frequency and intensity. The amplitude experiment indicates that discrimination is not sensitive to amplitude, concurring with earlier psychophysical results⁷. Second, additional psychophysical tests showed that animals truly discriminated between stimuli; that is, they compared the base and comparison stimuli (as opposed to, for example, classifying one of them)⁷. In all experiments, the base frequency changed unpredictably from trial to trial. This is the key: a given artificial signal led to different outcomes, depending only on the frequency of the mechanical base stimulus at any trial. Third, animals continuously switched between purely mechanical and microstimulation conditions with almost identical performance levels. Such high accuracy, based on the interaction between natural and artificially evoked activity, is consistent with the induction of a sensory percept^{13,14}. Such a causal link between perception and neuronal activity has rarely been quantified. Previous studies showed that a monkey's decision can be artificially biased in a visual motion-

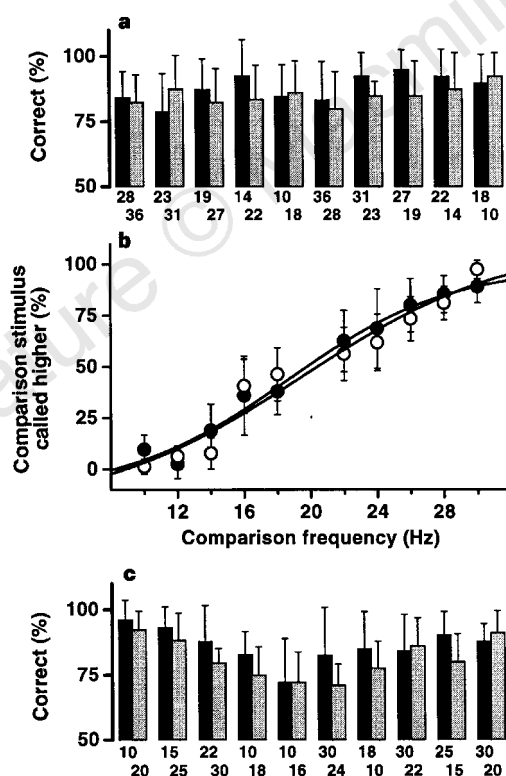


Figure 3 Psychophysical performance using aperiodic comparison stimuli. We used the same frequency sets, labels and trials per point as in Fig. 2. Black bars and circles indicate trials in which the comparison stimulus was mechanical and aperiodic. Grey bars and open symbols indicate intracortical microstimulation trials, from different runs, in which the comparison was electrical and aperiodic. The base stimulus was always mechanical and periodic. **a**, Results from one stimulus set with ten frequency pairs. **b**, Data and sigmoidal fits for ten pairs in which the base frequency was 20 Hz. **c**, Results from ten other pairs randomly intermixed with those in **b**.

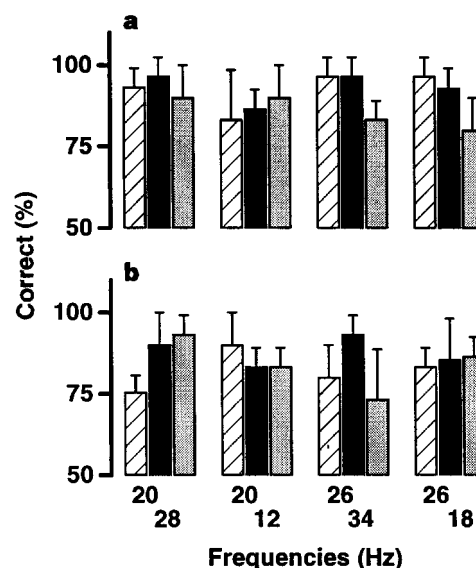


Figure 4 Psychophysical performance as a function of comparison amplitude. Upper and lower rows of numbers on the x axis indicate base and comparison frequencies, respectively, for each triplet of bars. **a**, Comparison stimulus was mechanical. **b**, Comparison stimulus was electrical. The amplitude of the comparison took one of three values: the standard amplitude A (black bars), $0.6A$ (hatched bars) or $1.4A$ (grey bars). All stimuli were periodic. Each data point includes 30 trials from three different runs. All conditions shown were randomly intermixed in each run.

discrimination paradigm^{13,14}, and that activity in a single afferent fibre can produce localized somatic sensations¹⁵. Here the evoked discharges from S1 provided the substrate for higher-order processes leading to frequency discrimination in those cortical areas that are anatomically linked to S1. Lesion studies indicate that S1 is necessary for the analysis of somatic stimuli^{16,17}. Thus, the microstimulation patterns used may elicit flutter sensations referred to the fingertips that are not unlike those felt with mechanical vibrations. □

Methods

Protocols. Recordings were obtained with an array of seven independent microelectrodes (refs 1, 18) (1–1.5 MΩ), one of which was used to microstimulate. Stimulation and recording sites changed from session to session. Recording procedures are described in ref. 1. In all paradigms, monkeys were rewarded for correct discrimination with a drop of liquid. Animals were handled according to institutional standards that meet or exceed those of the NIH and Society for Neuroscience.

Mechanical stimulation. Stimuli were delivered to the skin of the distal segment of digits 2, 3, or 4 of the left, restrained hand via a computer-controlled Chubbuck linear motor stimulator (2-mm round tip)¹⁷. Initial indentation was 500 μm. Stimulus amplitudes (*A*) were adjusted to equal subjective intensities¹⁷, for example, 71 μm at 12 Hz and 51 μm at 34 Hz (~1.4% per Hz).

Microstimulation. A computer-controlled pulse generator (Coulbourn), in series with an optical stimulus isolation unit, produced biphasic current pulses with the cathodal phase leading. Each phase lasted 0.2 ms, with 0.05 ms between phases. Two-pulse bursts, with 0.5 ms between pulses, were delivered at the comparison frequencies. Current amplitude varied between 65 μA and 100 μA. When compensations for subjective intensity were applied, amplitude was typically set at 95 μA for 12 Hz, with higher frequencies attenuated by ~1.5% per Hz.

Received 15 September; accepted 22 December 1997.

- Mountcastle, V. B., Steinmetz, M. A. & Romo, R. Frequency discrimination in the sense of flutter: psychophysical measurements correlated with postcentral events in behaving monkeys. *J. Neurosci.* **10**, 3032–3044 (1990).
- Talbot, W. H., Darian-Smith, I., Kornhuber, H. H. & Mountcastle, V. B. The sense of flutter vibration: comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey's hand. *J. Neurophysiol.* **31**, 301–334 (1968).
- Mountcastle, V. B., Talbot, W. H., Sakata, H. & Hyvärinen, J. Cortical neuronal mechanisms in flutter vibration studied in unanesthetized monkeys. *J. Neurophysiol.* **32**, 453–484 (1969).
- Kaas, J. H., Nelson, R. J., Sur, M. & Merzenich, M. M. Multiple representation of the body within the primary somatosensory cortex of primates. *Science* **204**, 511–512 (1979).
- Sur, M., Wall, J. T. & Kaas, J. H. Modular distribution of neurons with slowly adapting and rapidly adapting responses in area 3b of somatosensory cortex in monkeys. *J. Neurophysiol.* **51**, 724–744 (1984).
- Recanzone, G. H., Merzenich, M. M. & Schreiner, C. E. Changes in the distributed temporal response properties of SI cortical neurons reflect improvements in performance on a temporally based tactile discrimination task. *J. Neurophysiol.* **67**, 1071–1091 (1992).
- Hernández, A., Salinas, E., García, R. & Romo, R. Discrimination in the sense of flutter: new psychophysical measurements in monkeys. *J. Neurosci.* **17**, 6391–6400 (1997).
- Mountcastle, V. B. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.* **20**, 408–434 (1957).
- Powell, T. P. S. & Mountcastle, V. B. Some aspects of the functional organization of the cortex of the postcentral gyrus of the monkey: a correlation of findings obtained in a single unit analysis with cytoarchitecture. *Bull. Johns Hopkins Hosp.* **105**, 133–162 (1959).
- Siegel, S. & Castellan, N. J. *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York, 1988).
- LaMotte, R. H. & Mountcastle, V. B. The capacities of humans and monkeys to discriminate between vibratory stimuli of different frequency and amplitude: a correlation between neural events and psychophysical measurements. *J. Neurophysiol.* **38**, 539–559 (1975).
- Bair, W., Koch, C., Newsome, W. & Britten, K. Power spectrum analysis of bursting cells in area MT in the behaving monkey. *J. Neurosci.* **14**, 2870–2892 (1994).
- Salzman, C. D., Britten, K. H. & Newsome, W. T. Cortical microstimulation influences perceptual judgments of motion direction. *Nature* **346**, 174–177 (1990).
- Salzman, C. D., Murasugi, C. M., Britten, K. H. & Newsome, W. T. Microstimulation in visual area MT: effects on direction discrimination performance. *J. Neurosci.* **12**, 2331–2355 (1992).
- Vallbo, A. B. in *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 237–252 (MIT Press, Cambridge, Mass, 1995).
- LaMotte, R. H. & Mountcastle, V. B. Disorders in somesthesia following lesions of parietal lobe. *J. Neurophysiol.* **42**, 400–419 (1979).
- Zainos, A., Merchant, H., Hernández, A., Salinas, E. & Romo, R. Role of primary somatic sensory cortex in the categorization of tactile stimuli: effects of lesions. *Exp. Brain Res.* **115**, 357–360 (1997).
- Mountcastle, V. B., Reitschek, H. J., Poggio, G. F. & Steinmetz, M. A. Adaptation of the Reitschek method of multiple microelectrode recording to the neocortex of the waking monkey. *J. Neurosci. Methods* **36**, 77–84 (1991).

Acknowledgements. We thank W. T. Newsome for comments and discussions. R.R.'s research was partially supported by an International Research Scholars Award from the HHMI and by grants from DGAPA-UNAM, CONACyT and Fundación Miguel Alemán AC.

Correspondence and requests for materials should be addressed to R.R. (e-mail: rromo@ifscun1.ifisiol.unam.mx).

Primary afferent tachykinins are required to experience moderate to intense pain

Yu Qing Cao*, Patrick W. Mantyh†, Elaine J. Carlson‡, Anne-Marie Gillespie‡, Charles J. Epstein‡ & Allan I. Basbaum*

* Departments of Anatomy, Physiology and W. M. Keck Foundation Center for Integrative Neuroscience, †Department of Pediatrics, University of California San Francisco, UCSF Box 0452, San Francisco, California 94143-0452, USA

‡ Molecular Neurobiology Laboratory, Veterans Administration Medical Center and Department of Psychiatry, University of Minnesota, Minneapolis, Minneapolis 55417, USA

The excitatory neurotransmitter glutamate coexists with the peptide known as substance P in primary afferents that respond to painful stimulation¹. Because blockers of glutamate receptors reliably reduce pain behaviour^{2–4}, it is assumed that 'pain' messages are mediated by glutamate action on dorsal horn neurons. The contribution of substance P, however, is still unclear. We have now disrupted the mouse preprotachykinin A gene (*PPT-A*), which encodes substance P and a related tachykinin, neurokinin A (ref. 5). We find that although the behavioural response to mildly painful stimuli is intact in these mice, the response to moderate to intense pain is significantly reduced. Neurogenic inflammation, which results from peripheral release of substance P and neurokinin A (ref. 6), is almost absent in the mutant mice. We conclude that the release of tachykinins from primary afferent pain-sensing receptors (nociceptors) is required to produce moderate to intense pain.

Although substance P (SP) and neurokinin A (NKA) are synthesized together, most studies have focused on the contribution of SP. Rather than being necessary for the experience of pain, SP is considered to be a neuromodulator that alters the excitability of dorsal horn 'pain'-responsive neurons⁷. Other findings, however, indicate that SP may come into play under conditions that are different from those that evoke the release of glutamate exclusively. For example, higher frequencies of primary afferent stimulation are required to evoke the release of SP compared to glutamate^{8,9}, and we have shown in dorsal horn neurons that noxious stimulus-evoked internalization of the neurokinin-1 (NK-1) receptor, a marker of SP release, occurs in the normal animal only when stimulus intensities are well above the threshold for pain^{10,11}. On this basis,

Table 1 Density of neurokinin-receptor binding sites in *PPT-A*^{+/+}, *PPT-A*^{+/-} and *PPT*^{-/-} mice

		Cerebral cortex	Striatum	Spinal cord	Ileum
NK-1	WT	100 ± 18	100 ± 27	100 ± 22	100 ± 20
	HZ	93 ± 20	102 ± 4	109 ± 13	88 ± 16
	KO	103 ± 27	95 ± 10	96 ± 19	91 ± 25
NK-2	WT	–	–	–	100 ± 21
	HZ	–	–	–	124 ± 18
	KO	–	–	–	106 ± 24
NK-3	WT	100 ± 17	100 ± 18	100 ± 27	–
	HZ	74 ± 21	87 ± 16	115 ± 14	–
	KO	94 ± 14	111 ± 25	89 ± 23	–

Wild-type, WT; heterozygote, HZ; and homozygous mutants, KO. The density of receptor binding sites in WT animals was assigned a value of 100 ± s.e.m. and the densities of binding in HZ and KO animals were compared to this value. There were no statistically significant differences (ANOVA; *P* > 0.05) in the density of NK-1, NK-2 or NK-3 binding sites among the WT, HZ and KO animals in any of the four brain areas tested or in the ileum. Measurements were made from lamina I–V of the spinal cord dorsal horn and in the circular and longitudinal muscles of the ileum. Dashes indicate that no specific binding sites could be detected.

we considered that whereas glutamate signals all intensities of pain, SP could be required for the signalling of moderate to intense pain.

We therefore studied a variety of pain behaviours in mice with a targeted deletion of SP and NKA (Fig. 1). Homozygous progeny had the expected mendelian frequency and were fertile. Litter size and maternal behaviour were normal, as were performance on a rotating rod and activity in an open field. We did not detect SP or NKA immunoreactivity in the mutant mice (Fig. 1e), but autoradiographic mapping revealed normal levels of the three major tachykinin receptors (NK-1, 2 and 3) in the central nervous system (CNS)¹² and in peripheral tissue (ileum; Table 1). Moreover, the immunostaining pattern for the NK-1 receptor was not altered in the mutant mice, and the pattern and density of dorsal horn and/or dorsal root ganglion staining for the peptides, calcitonin-gene-related peptide, dynorphin, galanin, neuropeptide Y and somatostatin were unchanged. Noxious stimulation induced massive internalization of the NK-1 receptor in dorsal horn neurons of wild-type but not mutant mice.

The latency to respond to a mildly painful stimulus (a 52.5 °C hot plate) did not differ in the wild-type and mutant mice (Fig. 2a). When the plate temperature was increased to 55.5 °C, the latency to

respond decreased in the wild-type but not in the mutant mice. At a higher temperature (58.5 °C), the latencies in the two groups were again the same. Because the hot plate evokes behaviour that is integrated at the spinal and supraspinal levels, it is difficult to specify the locus of the contribution of SP and NKA. Comparable results, however, were observed in a test that monitors reflex withdrawal to noxious heating of the hindpaw (Fig. 2b). Taken together, these results indicate that primary afferent-derived tachykinins are required for normal pain responsiveness, but only to a 'window' of noxious thermal stimulus intensities.

Although the threshold for evoking a withdrawal reflex of the paw to a mechanical stimulus did not differ in wild type and mutant mice (Fig. 2c), the latency to respond to a more intense mechanical stimulus (tail clip) was significantly prolonged in the mutant mice (Fig. 2d). The pain behaviour (licking) evoked by a hindpaw injection of capsaicin, an intensely noxious chemical stimulus that directly activates C-fibres, was also significantly reduced in the mutant mice (Fig. 3a). The behaviour of the heterozygote in response to capsaicin did not differ from the wild type (see below). We also found decreased pain behaviour in the first phase of the formalin test, which provides a different measure of the acute

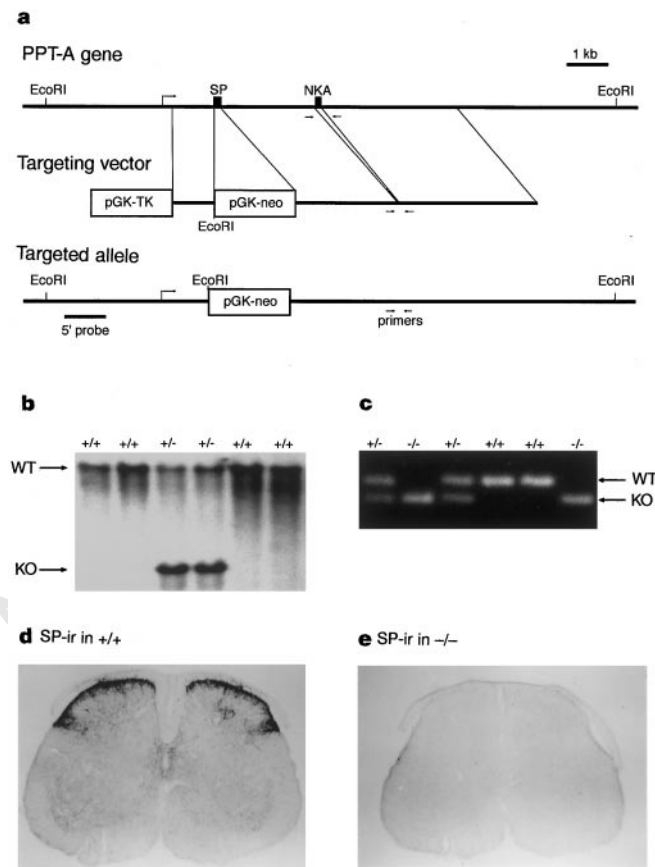


Figure 1 Generation of SP and NKA-deficient mice. **a**, *PPT-A* genomic locus, targeting vector and targeted allele. After homologous recombination, pGK-neo replaces the SP coding region and the NKA coding region is deleted. **b**, Southern blot of genomic DNA from targeted (+/-) and wild-type (+/+) embryonic stem cell clones, digested with *EcoRI* and probed with a 5' fragment outside the targeting vector. The position of the targeted (KO) and wild-type alleles (WT) are shown. **c**, PCR of tail DNA from one litter of a heterozygous crossing, using primers that flank the NKA coding region. The position of the targeted and wild-type alleles are shown. **d, e**, SP immunoreactivity (SP-ir) of lumbar spinal cord section from wild-type (**d**) and homozygous mutant (**e**) mice stained with anti-SP antibody. In the wild type, staining is concentrated in the superficial laminae of the dorsal horn; there is no staining in the knockout.

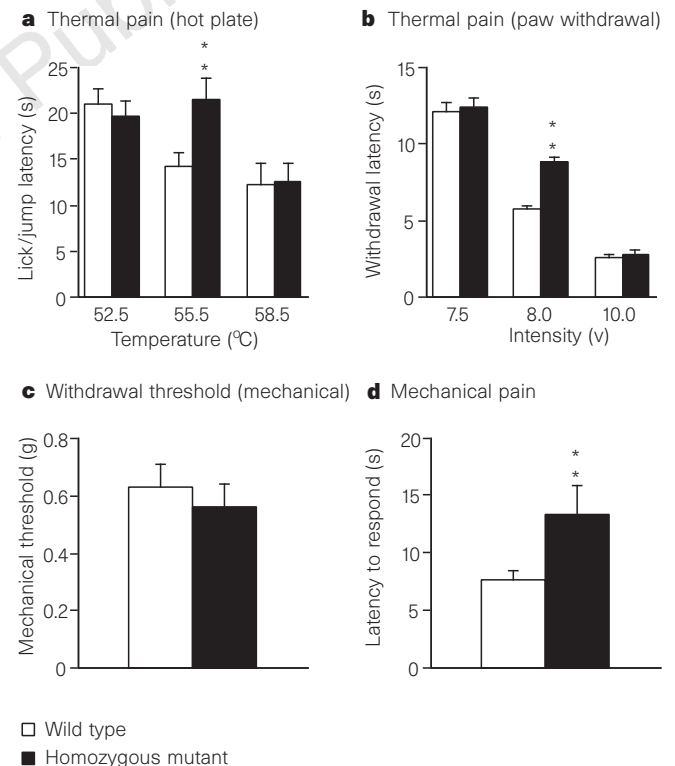


Figure 2 'Pain' responses to different thermal or mechanical stimulus intensities. **a**, Licking/jump latency in the hot-plate assay ($n = 12$); at 55.5 °C the homozygous mutant mice (black bars) showed a decreased pain response compared to wild type (white bars). **b**, Paw withdrawal latencies to noxious thermal stimuli ($n = 8$): at midrange intensities (8.0 volts), mutant mice also showed a decreased pain response. **c**, Withdrawal threshold to a mechanical stimulus (von Frey hair; $n = 16$); wild-type and mutant mice do not differ. **d**, Response latency to tail clip (wild type: $n = 31$; homozygous mutant: $n = 21$); the pain response of the mutant mice is significantly delayed. **, $P < 0.01$.

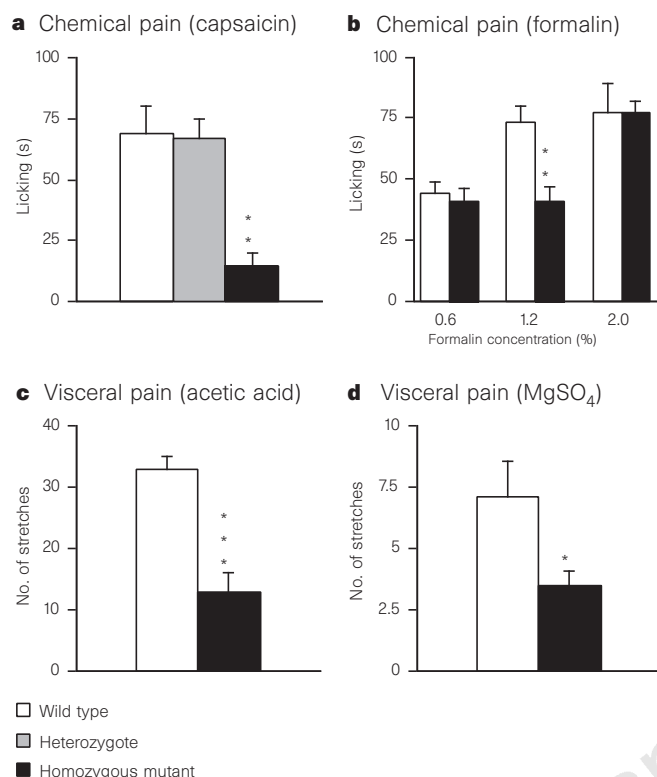


Figure 3 'Pain' responses to noxious chemical stimulation of cutaneous or visceral tissue. **a**, Duration of licking in response to intraplantar injection of capsaicin (4.5 mg; $n = 8$) is significantly reduced in mutant mice (black bars) compared to wild type (white bars) and heterozygous mice (grey bar); the last two groups do not differ. **b**, Duration of licking during phase I of the formalin test ($n = 8$; 0–10 min); only after an injection of 1.2% formalin is the reduced response of the mutant mice apparent. **c, d**, Visceral pain response (abdominal stretching) produced by intraperitoneal (i.p.) injection of dilute acetic acid (**c**), a stimulus associated with inflammation ($n = 12$ for both wild type and mutant) or i.p. injection of MgSO₄ (**d**), a stimulus that produces pain without inflammation (wild type: $n = 19$; homozygous mutant: $n = 16$). In both tests, the mutant mice showed significantly decreased responses. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

pain produced by direct chemical activation of C-fibres (Fig. 3b). Just as we observed for noxious heat, the intensity of the noxious chemical stimulus proved critical to revealing the contribution of SP/NKA. Thus, first-phase pain behaviour in response to 0.6% formalin was comparable in wild-type and mutant mice, but increasing the concentration to 1.2% increased the pain behaviour only in wild-type mice. With 2.0% formalin, the magnitude of pain behaviour in wild-type and mutant mice did not differ. Regardless of formalin concentration, we found no difference in pain behaviour in the second phase, which is proposed to result from a central sensitization process set up by activity during the first phase. This argues against an obligatory contribution of tachykinins to central sensitization. Finally, we tested the mice in two models of acute visceral pain, one (MgSO₄) that induces an immediate pain response which is independent of inflammation¹³, and one that is secondary to a delayed inflammatory response (intraperitoneal acetic acid). In both cases, we found reduced pain behaviour in the mutant mice (Fig. 3c, d).

These results indicate that tachykinins are released by polymodal nociceptive afferents¹⁴ and are critical neurotransmitters for the transmission and production of moderate to intense pain messages across modalities and from both somatic and visceral tissues. It follows from this that the dose of morphine required to produce analgesia would be lower in the knockout mice, but only at critical stimulus intensities. Consistent with this, we found that the half-maximal effective dose (ED₅₀) for morphine analgesia in the 55 °C hot-plate test was significantly lower in the mutant (ED₅₀ = 1.5 mg kg⁻¹) than in the wild-type mice (ED₅₀ = 3.9 mg kg⁻¹), but there was no difference when the stimulus was 52.5 °C (mutant ED₅₀ = 3.3 mg kg⁻¹, compared with 3.9 mg kg⁻¹ for wild type).

The importance of SP and NKA in neurogenic inflammation⁶ was also evident in the mutant mice. Specifically, we found a profound reduction of plasma extravasation, namely the leakage of protein from postcapillary venules, following topical application of capsaicin to the ear (Fig. 4a). The oedema produced by hindpaw injection of capsaicin was also reduced in the mutant mice (Fig. 4b). The magnitude of pain behaviour produced by oedema resulting from a similar injection was reduced in the heterozygote (Fig. 4b), as was the plasma extravasation after application of capsaicin to the ear (Fig. 4a). By contrast, the inflammation produced by complete Freund's adjuvant (CFA), which is non-neurogenic so does not

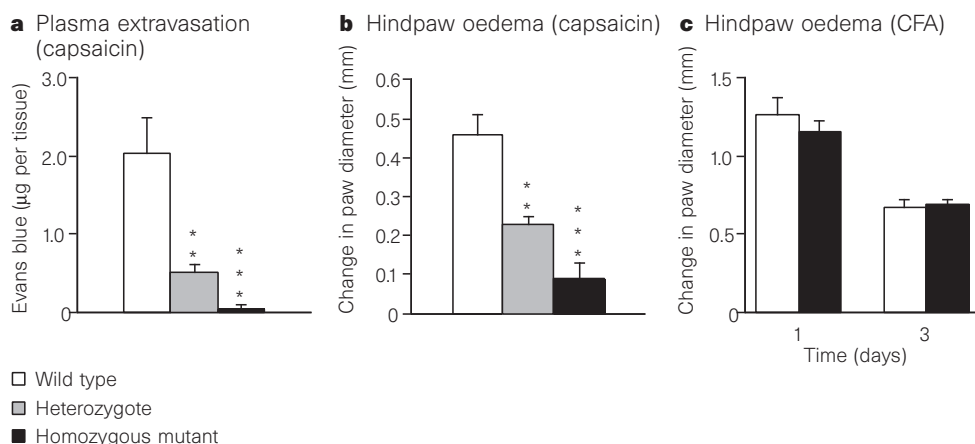


Figure 4 Magnitude of neurogenic and non-neurogenic inflammatory responses in wild-type (white bars), homozygous mutant (black bars) and heterozygous (grey bars) mice. **a**, Plasma extravasation induced by topical application of capsaicin on the ear ($n = 9$) is significantly reduced in the mutant mice, as is paw oedema (**b**) measured 30 min after intraplantar injection of capsaicin ($n = 8$). The

magnitude of these neurogenic inflammatory responses in the heterozygous mice are between those of the wild type and the mutant. **c**, Paw oedema 1 and 3 days after intraplantar injection of CFA ($n = 6$). Here the inflammation is non-neurogenic and there is no difference between the wild-type and mutant mice. **, $P < 0.01$; ***, $P < 0.001$.

depend on the integrity of primary afferent fibres¹⁵, was the same in the wild-type and mutant mice (Fig. 4c). These results indicate that there is a dissociation in the central and peripheral contribution of tachykinins that derive from primary afferent C-fibres. Neurogenic inflammation is dose-dependently related to the peripheral levels of SP and/or NKA, but a minimal content/release of these peptides from the central terminal of the afferent preserves the wild-type pain phenotype.

These studies all evaluated the response to acutely painful stimuli. Because clinical pain conditions that occur in the setting of persistent injury are characterized by significant decreases in the threshold for evoking pain, we also studied mechanical and thermal nociceptive thresholds in a CFA-induced model of inflammatory/nociceptive pain and in a nerve-injury-induced model of neuropathic pain¹⁶. We found that the allodynia (decreased nociceptive threshold) was the same in the wild-type and mutant mice (data not shown; $P > 0.05$), indicating that the nerve-growth-factor-mediated increase in SP that occurs in a CFA-treated animal¹⁷ is not necessary for the development and maintenance of allodynia, and that injury-induced central sensitization can occur in the absence of SP.

We have shown that mice carrying a disruption of the gene encoding SP/NKA have significantly reduced nociceptive pain responses when the stimuli are moderate to intense. The contribution of SP/NKA is neither modality- nor tissue-specific; pain behaviours evoked by thermal, mechanical and chemical stimulation of somatic and visceral tissues are all reduced in the mutant mice. The behavioural results suggest that glutamate is the relevant neurotransmitter when the intensity threshold for pain is crossed and that higher stimulus intensities evoke the release of SP/NKA; at the highest intensities, as-yet undefined systems must come into play. Because of the inconsistency in the antinociceptive effects of NK-1 receptor antagonists^{6,18}, it is unclear whether the deficit that we observed resulted from loss of a SP/NKA action at the NK-1 receptor, or whether the NK-2 and NK-3 receptors, both of which have been implicated in nociceptive processing^{19,20}, are involved. Differences in test paradigms, and the fact that antagonists are rarely completely selective for a particular receptor, makes it difficult to compare the studies. A loss of input to the NK-2 and NK-3 receptors could contribute to differences between the SP/NKA mutant phenotype and that of mice with a deletion of the NK-1 receptor^{21,22}. Finally, the results we obtained by using morphine in the mutant mice indicate that combination of an appropriate neurokinin receptor antagonist and morphine may prove effective for the relief of moderate to intense pain, with fewer side effects, using lower doses of morphine that are typically required when opioids are used alone.

Note added in proof: A. Zimmer *et al.*²⁹ recently generated mice with a deletion of SP and NKA and also found reduced pain responses; however, there are significant differences in the phenotype. Whether these differences reflect genetic background or experimental approach remains to be determined. □

Methods

Gene targeting. The mouse *PPT-A* gene was cloned by long polymerase chain reaction (PCR) from C57BL/6 mouse DNA²³. To target the deletion of SP and NKA, we replaced the SP-coding region in exon 3 of the *PPT-A* gene with *PGK-neo* and deleted the NKA-coding region in exon 6. The targeting vector was derived from pNTK by introducing a 1.1-kb genomic fragment 5' of the SP coding region upstream, and a 6.2-kb fragment 3' of the SP coding region downstream of *PGK-neo*. Homologous recombination in ES cell line CB1-4 (ref. 24) was confirmed by Southern blotting after *EcoRI* digestion and hybridization with 5', 3' and neo probes. Two positive clones microinjected into CD-1 morula-stage zygotes gave rise to chimaeric offspring, which were crossed to CD-1 mice to produce heterozygotes. Non-sibling heterozygotes were crossed to generate wild-type (+/+), heterozygous (+/-) and knockout (-/-) mice.

Immunocytochemistry. We immunostained frozen sections of brain and

spinal cord cut transversely at 30 μ m and dorsal root ganglia cut at 10 μ m on a cryostat using antisera directed against SP, calcitonin-gene-related peptide, dynorphin, galanin, neuropeptide Y, somatostatin (Peninsula Labs; 1:50,000) or the NK-1 receptor²⁵ (1:20,000) as described¹⁰.

Autoradiography and receptor binding. Tissues were cut transversely at 15 μ m on a cryostat and bound with a 100 pM solution of ¹²⁵I-labelled SP (Amersham), ¹²⁵I-labelled NKA (New England Nuclear) or ¹²⁵I-labelled NKB (New England Nuclear). Autoradiography has been described²⁶. In all cases, the density of binding of the wild type is presented as $100 \pm$ s.e.m. and the density of binding for the heterozygous and mutant mice was compared to this value.

General motor function. Two- and 6-month-old mice of both sexes were observed in an illuminated, sound-attenuated open field (30 $\times$ 30 cm, with lines dividing the floor into 3 $\times$ 3 squares) for 10 min. We counted the number of squares and centre squares crossed. One-, 3- and 6-month-old mice of both sexes were also tested on a rotating rod after three training sessions. We measured the average time (3 trials) that the mice stayed on the rod as it accelerated from 4 to 40 r.p.m. over a 5-min period.

Responses to noxious stimuli. Mice of both sexes were used (9 to 15 weeks old). Each mouse was acclimatized in the test apparatus for 30–60 min. All experiments were done blind to the genotype of the mice and data were analysed by ANOVA with Fisher's PLD post-hoc test. (1) Thermal stimuli. We recorded the latency period before the animal licked its hindpaw or jumped on the hot plate. The cut-off time was 60 s for the 52.5°C hot plate and 30 s at 55.5°C and 58.5°C. We also measured thermal paw withdrawal latencies to radiant heat²⁷. (2) Mechanical stimuli. We determined the withdrawal threshold of the hindlimb to a mechanical stimulus using calibrated von Frey hairs²⁸ and the latency to respond (lick or bite) to an intense mechanical stimulus (a clip applied to the tail). (3) Chemical stimuli. We evaluated the time spent licking the hindpaw after it was injected (maximum, 5 min). Paw oedema was measured with a spring-loaded caliper 30 min after the injection. For the formalin test, we injected 10 μ l dilute formalin (0.6, 1.25 or 2.0%) into the right hindpaw and monitored the total licking time during phase 1 (0 to 10 min after injection) and phase 2 (10 to 45 min after injection). (4) Visceral stimuli. We counted the number of abdominal stretches that occurred within 20 min after i.p. injection of 5.0 ml kg⁻¹ 0.6% acetic acid, a stimulus that produces visceral pain with inflammation, or within 5 min after i.p. injection of 120 mg kg⁻¹ MgSO₄ which produces visceral pain without inflammation.

Plasma extravasation. Anaesthetized mice (ketamine: 50 mg kg⁻¹; xylazine: 10 mg kg⁻¹) were injected with 30 mg kg⁻¹ Evan's blue through the tail vein. 5 min later, we painted both sides of the right ear with a solution of 250 μ g capsaicin in acetone. 30 min later, we extracted the extravasated Evan's blue from the ear and measured its concentration spectrophotometrically at 620 nm. The contralateral, control ear was painted with acetone vehicle.

Persistent-injury models. Tissue injury was produced by intraplantar injection of 10 μ g CFA in 20 μ l mineral oil (50%). To reproduce nerve injury, we ligated $\leq 1/3$ to $1/2$ of the sciatic nerve in the mid thigh. One and 3 days after the CFA and before and up to 2 weeks after nerve injury, we measured the thermal and mechanical paw-withdrawal thresholds using radiant heat and von Frey hairs, respectively. We also measured paw oedema produced by CFA.

Received 1 December 1997; accepted 17 February 1998.

- De Biasi, S. & Rustioni, A. Glutamate and substance P coexist in primary afferent terminals in the superficial laminae of spinal cord. *Proc. Natl Acad. Sci. USA* **85**, 7820–7824 (1988).
- Meller, S. T., Dykstra, C. L. & Gebhart, G. F. Acute mechanical hyperalgesia is produced by coactivation of AMPA and metabotropic glutamate receptors. *Neuroreport* **4**, 879–892 (1993).
- Wilcox, G. L. in *Proceedings of the Vth World Congress on Pain* (eds Bond, M. R., Charlton, J. E. & Woolf, C. J.) 97–111 (Elsevier Science, New York, 1991).
- Hunter, J. C. & Singh, L. Role of excitatory amino acid receptors in the mediation of the nociceptive response to formalin in the rat. *Neurosci. Lett.* **174**, 217–221 (1994).
- Nawa, H., Kotani, H. & Nakanishi, S. Tissue-specific generation of two preprotachykinin mRNAs from one gene by alternative RNA splicing. *Nature* **312**, 729–734 (1984).
- Amann, R., Schuligoi, R., Holzer, P. & Donnerer, J. The non-peptide NK1 receptor antagonist SR140333 produces long-lasting inhibition of neurogenic inflammation, but does not influence acute chemo- or thermoreception in rats. *Naunyn-Schmied. Arch. Pharmacol.* **352**, 201–205 (1995).
- Rusin, K. I., Ryu, P. D. & Randic, M. Modulation of excitatory amino acid responses in rat dorsal horn neurons by tachykinins. *J. Neurophysiol.* **68**, 265–286 (1992).
- Marvizon, J. C. G., Martinez, V., Grady, E. F., Bunnett, N. W. & Mayer, E. A. Neurokinin 1 receptor internalization in spinal cord slices induced by dorsal root stimulation is mediated by NMDA receptors. *J. Neurosci.* **17**, 8129–8136 (1997).
- Duggan, A. W., Riley, R. C., Mark, M. A., MacMillan, S. J. A. & Schaible, H. G. Afferent volley patterns and the spinal release of immunoreactive substance P in the dorsal horn of the anaesthetized spinal cat. *Neuroscience* **65**, 849–858 (1995).

10. Abbadié, C., Trafton, J., Liu, H., Mantyh, P. W. & Basbaum, A. I. Inflammation increases the distribution of dorsal horn neurons that internalize the neurokinin-1 receptor in response to noxious and non-noxious stimulation. *J. Neurosci.* **17**, 8049–8060 (1997).
11. Mantyh, P. W. *et al.* Receptor endocytosis and dendrite reshaping in spinal neurons after somatosensory stimulation. *Science* **268**, 1629–1632 (1995).
12. Maggi, C. A. The mammalian tachykinin receptors. *Gen. Pharmacol.* **26**, 911–944 (1995).
13. Gyires, K. & Torma, Z. The use of the writhing test in mice for screening different types of analgesics. *Arch. Int. Pharmacodyn.* **267**, 131–140 (1984).
14. Lawson, S. N., Crepps, B. A. & Perl, E. R. Relationship of substance P to afferent characteristics of dorsal root ganglion neurons in guinea-pig. *J. Physiol. (Lond.)* **505**, 177–191 (1997).
15. Hylden, J. L., Noguchi, K. & Ruda, M. A. Neonatal capsaicin treatment attenuates spinal Fos activation and dynorphin gene expression following peripheral tissue inflammation and hyperalgesia. *J. Neurosci.* **12**, 1716–1725 (1992).
16. Malmberg, A. B., Chen, C., Tonegawa, S. & Basbaum, A. I. Preserved acute pain and reduced neuropathic pain in mice lacking PKC γ . *Science* **278**, 279–283 (1997).
17. Safieh-Garabedian, B., Poole, S., Allchorne, A., Winter, J. & Woolf, C. J. Contribution of interleukin-1 beta to the inflammation-induced increase in nerve growth factor levels and inflammatory hyperalgesia. *Br. J. Pharmacol.* **115**, 1265–1275 (1995).
18. Rupniak, N. M. J., Carlson, E., Boyce, S., Webb, J. K. & Hill, R. G. Enantioselective inhibition of the formalin paw late phase by the NK1 receptor antagonist L-733,060 in gerbils. *Pain* **67**, 189–195 (1996).
19. Laneuville, O., Dorais, J. & Couture, R. Characterization of the effects produced by neurokinins and three agonists selective for neurokinin receptor subtypes in a spinal nociceptive reflex of the rat. *Life Sci.* **42**, 1295–1305 (1988).
20. Munro, F. E., Fleetwood-Walker, S. M., Parker, R. M. & Mitchell, R. The effects of neurokinin receptor antagonists on mustard oil-evoked activation of rat dorsal horn neurons. *Neuropeptides* **25**, 299–305 (1993).
21. Herrero, J. F. *et al.* Lack of “wind-up” of somatic nociceptive reflexes and persistence of visceral nociception in a NK1 receptor knock-out mouse. *Neurosci. Abstr.* **23**, 2354 (1997).
22. De Felipe, C. *et al.* Characterization of the NK-1 receptor knockout mouse: alterations in nociceptive behavior. *Neurosci. Abstr.* **23**, 2354 (1997).
23. Kako, K., Munekata, E., Hosaka, M., Murakami, K. & Nakayama, K. Cloning and sequence analysis of mouse cDNAs encoding preprotachykinin A and B. *Biomed. Res.* **14**, 253–259 (1993).
24. Li, Y. *et al.* Dilated cardiomyopathy and neonatal lethality in mutant mice lacking manganese superoxide dismutase. *Nature Genet.* **11**, 376–381 (1995).
25. Vigna, S. R. *et al.* Characterization of antibodies to the rat substance P (NK-1) receptor and to a chimeric substance P receptor expressed in mammalian cells. *J. Neurosci.* **14**, 834–845 (1994).
26. Mantyh, P. W., Gates, T., Mantyh, C. R. & Maggio, J. E. Autoradiographic localization and characterization of tachykinin receptor binding sites in the rat brain and peripheral tissues. *J. Neurosci.* **9**, 258–279 (1989).
27. Hargreaves, K., Dubner, R., Brown, F., Flores, C. & Joris, J. A new and sensitive method for measuring thermal nociception in cutaneous hyperalgesia. *Pain* **32**, 77–88 (1988).
28. Chaplan, S. R., Bach, F. W., Pogrel, J. W., Chung, J. M. & Yaksh, T. L. Quantitative assessment of tactile allodynia in the rat paw. *J. Neurosci. Meth.* **53**, 55–63 (1994).
29. Zimmer, A. *et al.* Hypoalgesia in mice with a targeted deletion of the tachykinin 1 gene. *Proc. Natl Acad. Sci. (USA)* **95**, 2630–2635 (1998).

Acknowledgements. We thank A. Doupe, D. Julius, J. Levine, S. Lisberger, A. Malmberg and W. Martin for critical comments on the manuscript. R. Montenson for the pNTK vector and S. Vigna for antisera against the NK-1 receptor. This work was supported by the NINDS, NIDR, NIDK of the NIH and an RRP from HHMI.

Correspondence and requests for materials should be addressed to A.I.B. (e-mail: aib@phy.ucsf.edu).

Altered nociception, analgesia and aggression in mice lacking the receptor for substance P

Carmen De Felipe*, Juan F. Herrero†, John A. O'Brien‡, James A. Palmer‡, Christopher A. Doyle‡, Andrew J. H. Smith§, Jennifer M. A. Laird†, Carlos Belmonte*, Fernando Cervero† & Stephen P. Hunt‡

* Instituto de Neurociencias, Universidad Miguel Hernandez, Ap. correos 374, 03080 Alicante, Spain

† Departamento de Fisiología, Universidad de Alcalá, Facultad de Medicina, Campus Universitario, Alcalá de Henares, 28871 Madrid, Spain

‡ Neurobiology Division, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

§ Centre for Genome Research, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JQ, UK

The peptide neurotransmitter substance P modulates sensitivity to pain by activating the neurokinin-1 (NK-1) receptor, which is expressed by discrete populations of neurons throughout the central nervous system^{1–4}. Substance P is synthesized by small-diameter sensory ‘pain’ fibres⁵, and release of the peptide into the dorsal horn of the spinal cord following intense peripheral stimulation⁶ promotes central hyperexcitability and increased sensitivity to pain^{7–10}. However, despite the availability of specific

NK-1 antagonists⁴, the function of substance P in the perception of pain remains unclear. Here we investigate the effect of disrupting the gene encoding the NK-1 receptor in mice. We found that the mutant mice were healthy and fertile, but the characteristic amplification (‘wind up’) and intensity coding of nociceptive reflexes was absent. Although substance P did not mediate the signalling of acute pain or hyperalgesia, it was essential for the full development of stress-induced analgesia and for an aggressive response to territorial challenge, demonstrating that the peptide plays an unexpected role in the adaptive response to stress.

Homologous recombination in embryonic stem cells was used to create a 129/Sv × C57BL/6 mouse line in which the NK-1 receptor gene was disrupted in exon 1 (ref. 11; Fig. 1a). Interbreeding of mice heterozygous for the disrupted allele resulted in healthy homozygous mutant mice (Fig. 1b), which bred normally and did not show impaired maternal behaviour.

Receptor autoradiography with [¹²⁵I]Bolton-Hunter substance P (SP)¹² indicated that SP binding sites were absent in the brain and spinal cord of NK1^{−/−} mice (Fig. 1c, d), and therefore that no receptor was available to mediate the actions of SP in NK1^{−/−} mice. This was confirmed by immunohistochemistry using a specific antibody raised against the carboxy terminus of the NK-1 receptor (data not shown). No detectable immunohistochemical differences

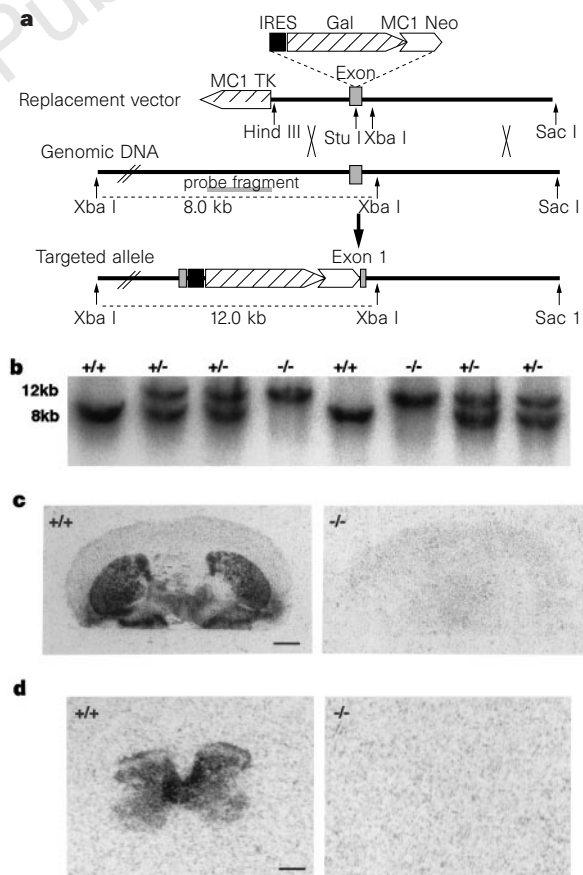


Figure 1 NK-1 receptor gene disruption by homologous recombination. **a**, Drawing of the region of the wild-type NK-1 locus containing exon 1, the targeting (replacement) vector and the predicted structure of the targeted NK-1 receptor gene. The 5' external probe used for Southern blot analysis and the sizes of the restriction fragments detected with this probe in the wild-type and targeted alleles are indicated. **b**, Southern blot analysis of offspring from an NK1^{+/−} intercross by *Xba*I digestions of tail DNA and hybridization with the above probe. Homozygous (12 kb), heterozygous (12 kb and 8 kb) and wild-type (8 kb) mice were generated at a ratio of 1:2:1. **c, d**, Autoradiographic mapping of SP binding sites in forebrain (**c**) and spinal cord (**d**) sections of NK1^{+/−} and NK1^{+/+} mice. Scale bars, 1 mm (**c**) and 0.5 mm (**d**).

were found in the intensity or distribution of SP or CGRP (calcitonin gene related peptide) staining in the spinal cord and brain of $NK1^{+/+}$ and $NK1^{-/-}$ mice. A count of β -galactosidase-positive floorplate cells and neurons in developing¹³ and adult $NK1^{-/-}$ animals (see Supplementary Information) revealed that no neuronal cell death had occurred as a result of gene manipulation (data not shown).

Electrophysiological evidence implicates the NK-1 receptor in spinal cord sensitization after injury or inflammation, and in the amplification of postsynaptic responses following repetitive high-threshold nerve stimulation⁷⁻⁹. We investigated these functions by recording electromyographic activity¹⁴ in the anaesthetized mouse while applying electrical or mechanical stimulation to the hind paw. In response to noxious mechanical stimulation at graded intensities, $NK1^{-/-}$ mice did not show intensity encoding. In $NK1^{+/+}$ mice, increasing the stimulus intensity progressively increased the response ($P < 0.05$; Fig. 2a). For example, a force of 100 mN above threshold produced an average increase of 101%, whereas a supra-threshold force of 300 mN induced an average increase of 381%. In contrast, $NK1^{-/-}$ mice showed no clear difference in their responses to stimuli 100–600 mN above threshold (Fig. 2b). Both the response variability and the firing rate in response to mechanical stimulation at threshold intensities were higher in the $NK1^{-/-}$ mice than in the $NK1^{+/+}$ mice ($P < 0.05$) (Fig. 2b). The higher firing rates in $NK1^{-/-}$ mice may be due to a reduced level of descending inhibitory control resulting from reduced nociceptive input. Reduction in inhibitory control has been demonstrated in adult rats treated at birth with capsaicin¹⁵, a neurotoxin that eliminates over 90% of sensory afferent unmyelinated C-fibres and therefore destroys most of the SP-containing sensory input.

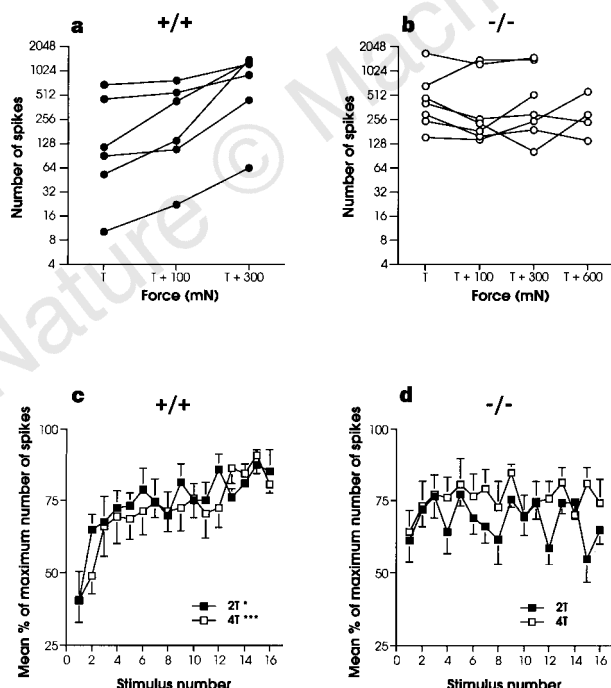


Figure 2 Nociceptive withdrawal reflex response following mechanical and electrical stimulation of the hind paw. **a, b**, Responses to noxious mechanical stimulation at different intensities. Note log scale and the encoding of the responses in the $NK1^{+/+}$ mice (**a**, $P < 0.05$ compared to threshold) but not in those from the $NK1^{-/-}$ mice (**b**). **c, d**, Electrical stimulation induces 'wind up' in $NK1^{+/+}$ mice (**c**, $n = 6$; $P < 0.001$ for 4T; T is threshold) but not in $NK1^{-/-}$ mice (**d**, $n = 7$; $P < 0.05$ for 2T). There was no difference in the baseline response between the two groups. The wind-up data were normalized to the maximal responses evoked. Data are mean $\pm$ s.e.m. and were analysed using the ANOVA for repeated measures with post-hoc Tukey's tests.

We also examined 'wind up', in which repetitive activation of unmyelinated C-fibres but not myelinated A-fibres increases the response to subsequent C-fibre input. This effect is reduced by NK1- and NMDA (*N*-methyl-D-aspartate)-receptor antagonists^{8,9,16}. In $NK1^{+/+}$ mice, the mean number of spikes increased by 217% with twice threshold stimulation, and by 224% with four-times threshold stimulation (Fig. 2c). In $NK1^{-/-}$ mice, wind up was not observed at any of the stimulation intensities used (Fig. 2d). We conclude that the NK-1 receptor is essential for the normal wind-up response.

Behaviourally, acute nociceptive thresholds were not affected by the gene disruption. Tail-pinch tests showed no significant differences in thresholds between the two groups of animals (Fig. 3a). In paw-withdrawal reflex experiments, the thresholds were again not significantly different between the two groups of animals, either for noxious mechanical or electrical stimulation (Fig. 3a). Tail-flick and hot-plate assays were used to assess heat thresholds. Tail flick is primarily a spinal reflex, whereas the hot-plate assay has a substantial supraspinal input to the response, which involves lifting and licking the hind paw^{17,18}. Latencies were comparable in both $NK1^{-/-}$ and $NK1^{+/+}$ animals (Fig. 3b, c). We conclude that substance P does not mediate acute pain sensation.

Peripheral release of substance P contributes to the neurogenic inflammatory response. The inflammatory response produced by capsaicin applied to the ear¹⁹ or by carrageenan injected into the hind paw was reduced by 74% and 62%, respectively, in $NK1^{-/-}$ mice compared to $NK1^{+/+}$ mice. The upregulation of SP expression in subsets of sensory and spinal cord neurons following inflammatory lesion of the hind paw has suggested that the increased release of SP may cause the central sensitization and hyperalgesia associated with inflammation^{10,20}. We investigated this possibility after injection of complete Freund's adjuvant into the left hind paw of mice. At 24 h or 60 h after injection of adjuvant, the inflammatory process was comparable in both groups of mice, with similar degrees of erythema and swelling of the ipsilateral paw ($NK1^{+/+}$, 3.48 ± 0.11 mm; $NK1^{-/-}$, 3.29 ± 0.09 mm; mean $\pm$ s.e.m.), indi-

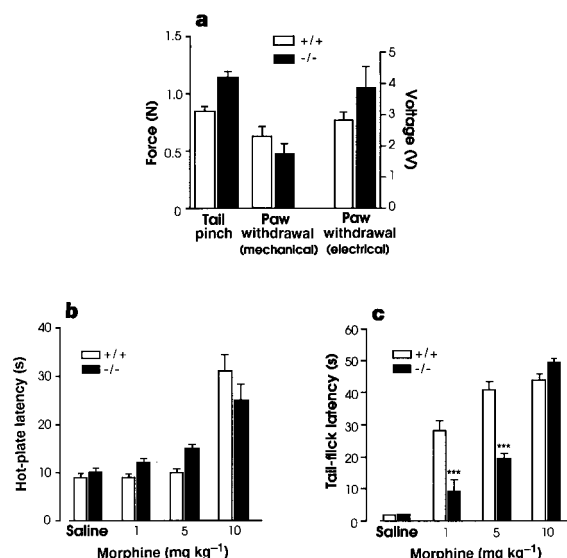


Figure 3 Acute nociceptive responses and sensitivity to morphine. **a**, Thresholds for withdrawal from tail pinch in awake animals and reflex threshold to noxious mechanical (force) or electrical (voltage) stimulation of the paw in anaesthetized animals ($NK1^{+/+}$, $n = 6$; $NK1^{-/-}$, $n = 7$). **b, c**, The effects of intraperitoneal morphine sulphate on hot-plate (**b**) and tail-flick (**c**) latency ($NK1^{+/+}$, $n = 10$; $NK1^{-/-}$, $n = 10$). *** $P < 0.001$, compared to $NK1^{+/+}$. Data are mean $\pm$ s.e.m. and were analysed with the Mann-Whitney U-test.

cating that SP is not involved in the peripheral neurogenic inflammatory response to adjuvant. The development of ipsilateral mechanical hyperalgesia was similar in both groups of mice (Fig. 4a). However, there was significant contralateral hyperalgesia ($P < 0.01$) in $NK1^{-/-}$ mice but not $NK1^{+/+}$ mice (Fig. 4a), although without obvious swelling of the right paw ($NK1^{+/+}$, 2.16 ± 0.09 mm; $NK1^{-/-}$, 2.10 ± 0.07 mm; mean $\pm$ s.e.m.). These data indicate that hyperalgesia can develop in the absence of SP activity.

Nociceptive behaviour following formalin injection was also disrupted in the $NK1^{-/-}$ mice. The second phase of nociceptive behaviour^{21,22} was attenuated by 45% in mutant mice. Numbers of Fos-positive neurons²³ were reduced by 30% within the superficial laminae (I–II) of $NK1^{-/-}$ mice, as in antagonist-treated animals²⁴,

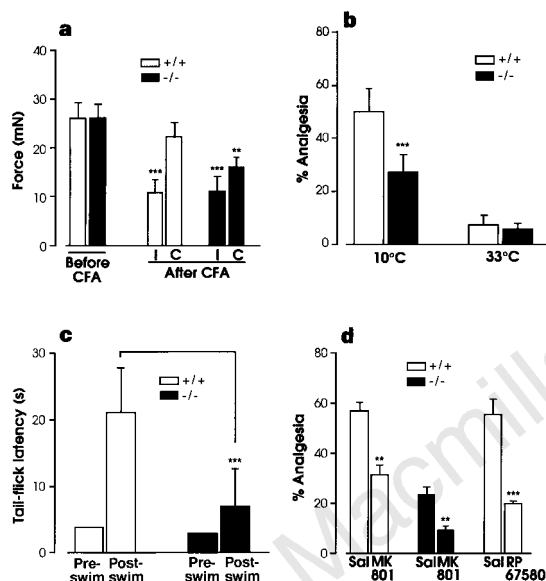


Figure 4 Adjuvant-induced inflammation and swim-stress-induced analgesia. **a**, Hyperalgesia measured with Von Frey hairs 60 h after ipsilateral (I) hind paw injection of complete Freund's adjuvant. C, contralateral. ($NK1^{+/+}$, $n = 12$; $NK1^{-/-}$, $n = 12$). $^{**}P < 0.01$ compared to $NK1^{+/+}$ before adjuvant; $^{***}P < 0.001$ compared to same phenotype before adjuvant, analysed by the Wilcoxon signed rank test. **b**, Percentage analgesia induced by a 3-min swim at 10°C or 33°C assessed using the hot-plate assay ($NK1^{+/+}$, $n = 16$; $NK1^{-/-}$, $n = 16$). $^{***}P < 0.001$ compared to 10°C swim $NK1^{+/+}$. **c**, Tail-flick latency measured following a 90-s swim in water at 4°C ($NK1^{+/+}$, $n = 10$; $NK1^{-/-}$, $n = 10$). $^{***}P < 0.001$ compared to post-swim $NK1^{+/+}$. **d**, Effect of MK-801 (0.075 mg kg⁻¹, i.p.) or RP 67580 (0.04 mg kg⁻¹, i.v.) on swim-stress-induced analgesia at 10°C using the hot-plate assay ($NK1^{+/+}$, $n = 6$; $NK1^{-/-}$, $n = 6$). Sal, saline. $^{**}P < 0.01$, $^{***}P < 0.001$ compared to saline. Data are mean $\pm$ s.e.m.

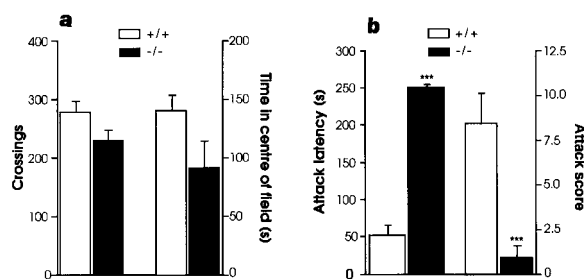


Figure 5 Anxiety and aggression. **a**, The open-field assay. Number of line crossings and time spent in centre field were measured ($NK1^{+/+}$, $n = 10$; $NK1^{-/-}$, $n = 10$). **b**, Aggressive behaviour of mice in the resident-intruder assay. Attack latencies and total fighting scores of $NK1^{+/+}$ ($n = 8$) and $NK1^{-/-}$ ($n = 8$) mice. $^{***}P < 0.001$ compared to $NK1^{+/+}$. Data are mean $\pm$ s.e.m. and were analysed by the Mann-Whitney U-test.

indicating that there was a reduction in nociceptive input to the spinal cord.

We next examined the analgesic response in mice that were exposed either to morphine or to a period of stress induced by a short forced swim. Morphine increased hot-plate thresholds comparably in both $NK1^{+/+}$ and $NK1^{-/-}$ animals (Fig. 3b), whereas the increase in tail-flick thresholds with low doses of morphine was reduced significantly in $NK1^{-/-}$ mice (Fig. 3c). Tail-flick and hot-plate responses are modulated by different opioid mechanisms within the central nervous system^{25–27}. We conclude that the modulation of the tail-flick spinal reflex by opiates is mediated in part by SP, as previously suggested²⁸.

Cold water (4–15°C) produces a non-opiate-, NMDA-dependent analgesia, whereas warmer water (33°C) results in opiate-dependent analgesia^{25,26}. Both types of analgesia are thought to be mediated by descending supraspinal pathways. After a swim at 4°C (tail-flick test) or 10°C (hot-plate test), the development of analgesia was reduced markedly in the $NK1^{-/-}$ mice compared to $NK1^{+/+}$ mice: a 45% reduction was seen in hot-plate latency (Fig. 4b), and a 70% reduction was seen in tail-flick latency (Fig. 4c). To confirm that this was not due to a neurochemical compensation for the absence of NK-1 receptors during development, we tested $NK1^{+/+}$ animals in the hot-plate assay after pretreatment with the NK1-specific antagonist RP67580. Antagonist-treated $NK1^{+/+}$ mice showed a reduction in analgesia comparable to that seen in $NK1^{-/-}$ mice (Fig. 4d), demonstrating that SP activates stress-induced analgesia. We also pretreated $NK1^{+/+}$ and $NK1^{-/-}$ mice with the NMDA-receptor-antagonist MK-801, which reduces stress-induced analgesia produced by cold-water swimming²⁶. Both groups of mice were equally sensitive to the antagonist in the hot-plate assay (Fig. 4d), and a comparable decline in analgesia was seen in both groups (~40%), indicating that the neural mechanisms that generate NMDA analgesia are probably intact in the $NK1^{-/-}$ mouse. Finally, mice were pretreated with the opiate antagonist naloxone, which abolishes the opiate-sensitive, stress-induced analgesia that follows a swim at 33°C. Both groups of animals showed similar analgesic effects and, following naloxone treatment, a comparable reduction in analgesia, indicating that opiate-mediated analgesia induced by stress was intact in $NK1^{-/-}$ mice (data not shown). Therefore, the generation of stress-induced analgesia by SP is pharmacologically distinct from the opiate-mediated pathway. We propose that SP controls spinal excitability primarily by activating central inhibitory pathways from the brainstem.

The NK-1 receptor is found throughout the brain, notably within the limbic system and hypothalamus^{1,2}, suggesting a role for SP in emotional behaviour. Anxiety was evaluated in the open-field assay¹⁷. No differences between groups were detected, and no preferences for the sides or the centre of the field were noticed (Fig. 5a), demonstrating that anxiety was similar in $NK1^{-/-}$ and $NK1^{+/+}$ animals. Aggression was assessed using the 'resident-intruder' test, in which isolated males of either phenotype are exposed to $NK1^{+/+}$ mice housed communally¹⁷. $NK1^{-/-}$ mice were considerably less aggressive (Fig. 5b). Consistent with these data, defensive rage provoked by medial hypothalamic stimulation is blocked by NK-1 antagonists²⁹.

We conclude that SP is important for orchestrating the response of the animal to major stressors such as pain, injury or invasion of territory. These biological responses have obvious survival value for the animal. In the wild, a short period of analgesia following injury facilitates the execution of a 'fight or flight' response, and the aggressive response secures the territorial rights and social status of the animal.

Methods

Targeting the NK1 receptor gene. A genomic DNA clone containing exon 1 of the mouse NK1 receptor gene was isolated by screening a λ 2001 mouse 129 library with a rat NK-1 cDNA probe³. A cassette containing an internal

ribosome entry site and the *lacZ* coding sequence¹¹, together with a neomycin-resistance gene expressed from its own promoter, was inserted into the unique *StuI* site in exon 1. For negative selection, two copies of an HSV-*tk* gene were inserted in the 5' polylinker¹¹. The vector was linearized and electroporated into HM1 ES cells. G418- and GANC-resistant clones were selected and identified. Two targeted clones were injected into C57BL/6 blastocysts, and chimaeric males were mated with C57BL/6 females. Transmission of the mutant allele was determined by Southern blot analysis of tail DNA. Mice homozygous for the NK1 mutation were produced by crossing heterozygotes.

Receptor binding. For NK1-receptor binding, fresh brain and spinal cord tissue was removed from adult mice and immediately frozen on dry ice. Cryostat sections (12 µm), collected on poly-L-lysine-coated slides, were incubated in 100 pM [¹²⁵I]Bolton-Hunter substance P (Amersham) as described¹². Control sections were incubated with excess (1 µM) [Sar⁹, Met(O₂)¹¹]substance P. Sections were dried overnight at room temperature and exposed to Hyperfilm-³H (Amersham) for up to 6 days.

Electrophysiology. Surgery was performed as described¹⁴, and withdrawal reflexes were elicited by noxious mechanical and electrical stimulation of the paw under α-chloralose anaesthesia. Threshold force and voltage were found by increasing pressure or voltage until an electrical response was recorded. Multi-unit electromyograph spikes composed of two or three units were recorded using a bipolar needle electrode inserted percutaneously into hind limb muscles. The protocol for stimulation was a 1-s pinch at threshold force, repeated 10 times, followed by a 1-min pause and then electrical stimulation of 16 pulses (2-ms pulse, 1 Hz) at C-fibre threshold voltage. The protocol was repeated at intervals of 2 min using stimulation intensities of 100–600 mN over mechanical threshold, and two and four times the threshold for electrical stimulation. (For details of electrophysiological methods, see Supplementary Information.) Peripheral inflammatory responses were measured as volume of the foot pad using a plethysmograph; increased ear thickness was measured with a micrometer. Capsaicin (10 µl of 25 mM) was applied to the ear for 30 min and the effect of carrageenan injection into the hind paw (20 µl, 10 mg per ml) was assessed after 6 h.

Behavioural studies. Animals (8–16 weeks) were housed in a temperature- and humidity-controlled environment with a 12-h light-and-dark cycle and free access to food and water. All testing was performed between 10:00 and 14:00 h. No differences were detected between male and female animals in the behavioural tasks (other than the 'resident-intruder' assay, for which only males were used).

Noxious mechanical threshold (tail pinch) was assessed by applying a 30-s ramp force (0 to 2 N) over a surface of 2 mm² situated 3 cm from the end of the tail. The force at which the animal moved or attempted to move the tail was considered as threshold force. The tail-flick assay consists of lightly restraining the mice and immersing two thirds of the tail in water at 52 °C. Time to the first movement of the tail was recorded. In the hot-plate assay, the mouse was placed on a hot plate at 52 °C and the time was recorded to the first hind-paw flick or lick (cut-off time, 60 s).

Stress-induced analgesia was assessed by first measuring the hot-plate or tail-flick latency and then placing the animal in a 4-litre beaker containing 2 litres of water. Optimum temperatures to produce analgesia differ between tasks^{25–27}. For the hot-plate assay, water was at either 10 °C or 33 °C, and the swim was for 3 min. For tail flick, water temperature was 4 °C, and the swim was for 90 s. Following the swim, animals were removed, patted dry with tissues and allowed to rest for either 2 min (following 3-min swim) or 10 min (following 90-s swim) before remeasurement heat thresholds. Morphine sulphate, MK-801 ((5R,10S)-(+)-5-methyl-10,11-dihydro-5H-dibenzo[a,d]cyclohepten-5,10-imine hydrogen maleate; RBI Inc.) and naloxone were administered intraperitoneally 20 min before testing; RP 67580 ((3aR,7aR)-2-[1-imino-2-(2-methoxyphenyl)ethyl]-7,7-diphenyl-4-perhydroisindolone; Rhône-Poulenc Rorer, Vitry-sur-Seine, France) was injected intravenously 20 min before measurements were taken. The percentage analgesia was calculated as described²⁶.

Formalin (25 µl; 2% paraformaldehyde in PBS) was injected into the plantar surface of the left hind paw. The time spent licking the injected paw was recorded as the nociceptive score. Inflammation was induced by injecting 10 µl of complete Freund's adjuvant (Difco) into the foot pad of the left hind paw under halothane anaesthesia. After 60 h, animals were placed on a raised grid, and mechanical thresholds were measured using Von Frey hairs.

The behaviour of mice was videotaped in an open field (50 × 50 cm) with lines dividing the floor area into 25 squares. Line crossing and the time spent in the 9 central squares was counted in 15-min blocks from the videotapes. Locomotor activity was measured over a 24-h period, using automated activity boxes.

For the resident male-intruder assay, resident males were housed individually for 28 d before the procedure. The intruder mice were housed in cages of four or five. The mice received one 5-min training session in the morning and two 5-min test sessions in the afternoon, all of which were videotaped. Each test session was divided into 10-s blocks. Every block in which an attack took place scored 1; blocks without an attack scored 0. All scores were added to give a final fighting score.

Received 23 July 1997; accepted 9 February 1998.

- Maeono, H., Kiyama, H. & Tohyama, M. Distribution of the substance P receptor (NK-1 receptor) in the central nervous system. *Mol. Brain Res.* **18**, 43–58 (1993).
- Nakaya, Y. *et al.* Immunohistochemical localization of substance P receptor in the central nervous system of the adult rat. *J. Comp. Neurol.* **347**, 249–274 (1994).
- Yokota, Y. *et al.* Molecular characterization of a functional cDNA for rat substance P receptor. *J. Biol. Chem.* **264**, 17649–17652 (1989).
- Maggi, C. A., Patacchini, R., Rovero, P. & Giachetti, A. Tachykinin receptors and tachykinin receptor antagonists. *J. Auton. Pharmacol.* **13**, 23–93 (1993).
- McCarthy, P. W. & Lawson, S. N. Cell type and conduction velocity of rat primary sensory neurons with substance P-like immunoreactivity. *Neuroscience* **28**, 745–753 (1989).
- Duggan, A. W., Hendry, I. A., Morton, C. R., Hutchison, W. D. & Zhao, Z. Q. Cutaneous stimuli releasing immunoreactive substance P in the dorsal horn of the cat. *Brain Res.* **451**, 261–273 (1988).
- Laird, J. M. A., Hargreaves, R. J. & Hill, R. G. Effect of RP 67580, a non-peptide neurokinin-1 receptor antagonist, on facilitation of a nociceptive spinal flexion reflex in the rat. *Br. J. Pharmacol.* **109**, 713–718 (1993).
- Ma, Q.-P. & Woolf, C. J. Involvement of neurokinin receptors in the induction but not the maintenance of mechanical allodynia in rat flexor motoneurons. *J. Physiol. (Lond.)* **486**, 769–777 (1995).
- Xu, X.-J., Dalsgaard, C.-J. & Wiesenfeld-Hallin, Z. Intrathecal CP-96,345 blocks reflex facilitation induced in rats by substance P and C-fiber-conditioning stimulation. *Eur. J. Pharmacol.* **216**, 337–344 (1992).
- Neumann, S., Doubell, T. P., Leslie, T. & Woolf, C. J. Inflammatory pain hypersensitivity mediated by phenotypic switch in myelinated primary sensory neurons. *Nature* **384**, 360–364 (1996).
- Nehls, M. *et al.* Two genetically separable steps in the differentiation of thymic epithelium. *Science* **272**, 886–889 (1996).
- Mantyh, P. W., Gates, T., Mantyh, C. R. & Maggio, J. E. Autoradiographic localization and characterization of tachykinin receptor binding sites in the rat brain and peripheral tissues. *J. Neurosci.* **9**, 258–279 (1989).
- De Felipe, C., Pinnock, R. D. & Hunt, S. P. Modulation of chemotaxis in the developing spinal cord by substance P. *Science* **267**, 899–902 (1995).
- Herrero, J. F. & Headley, P. M. Functional evidence for multiple receptor activation by κ-ligands in the inhibition of spinal nociceptive reflexes in the rat. *Br. J. Pharmacol.* **110**, 303–309 (1993).
- Cervero, F. & Plenderleith, M. B. C-fibre excitation and tonic descending inhibition of dorsal horn neurones in adult rats treated at birth with capsaicin. *J. Physiol. (Lond.)* **365**, 223–237 (1985).
- Woolf, C. J. & Thompson, S. W. N. The induction and maintenance of central sensitization is dependent on N-methyl-D-aspartic acid receptor activation; implications for the treatment of post-injury pain hypersensitivity states. *Pain* **44**, 293–299 (1991).
- König, M. *et al.* Pain responses, anxiety and aggression in mice deficient in pre-proenkephalin. *Nature* **383**, 535–538 (1996).
- Yaksh, T. L. & Malmberg, A. B. In *The Textbook of Pain* (eds Wall, P. D. & Melzack, R.) 165–200 (Churchill-Livingstone, London, 1994).
- Bozic, C. R., Lu, B., Höpken, U. E., Gerard, C. & Gerard, N. P. Neurogenic amplification of immune complex inflammation. *Science* **273**, 1722–1725 (1996).
- Noguchi, K. & Ruda, M. A. Gene regulation in an ascending nociceptive pathway: inflammation-induced increase in preprotachykinin mRNA in rat lamina I spinal projection neurons. *J. Neurosci.* **12**, 2563–2572 (1992).
- Abbadie, C., Taylor, B. K., Peterson, M. A. & Basbaum, A. I. Differential contribution of the two phases of the formalin test to the pattern of *c-fos* expression in the rat spinal cord: studies with remifentanyl and lidocaine. *Pain* **69**, 101–110 (1997).
- Rupniak, N. M. J., Carlson, E., Boyce, S., Webb, J. K. & Hill, R. G. Enantioselective inhibition of the formalin paw late phase by the NK₁ receptor antagonist L-733,060 in gerbils. *Pain* **67**, 189–195 (1996).
- Hunt, S. P., Pini, A. & Evan, G. Induction of *c-fos*-like protein in spinal cord neurons following sensory stimulation. *Nature* **328**, 632–634 (1987).
- Chapman, V., Buritova, J., Honoré, P. & Besson, J.-M. Physiological contributions of neurokinin 1 receptor activation, and interactions with NMDA receptors, to inflammatory-evoked spinal *c-Fos* expression. *J. Neurophysiol.* **76**, 1817–1827 (1996).
- Rubinstein, M. *et al.* Absence of opioid stress-induced analgesia in mice lacking β-endorphin by site-directed mutagenesis. *Proc. Natl Acad. Sci. USA* **93**, 3995–4000 (1996).
- Marek, P., Mogil, J. S., Sternberg, W. F., Panocka, I. & Liebeskind, J. C. N-Methyl-D-aspartic acid (NMDA) receptor antagonist MK-801 blocks non-opioid stress-induced analgesia. II. Comparison across three swim-stress paradigms in selectively bred mice. *Brain Res.* **578**, 197–202 (1992).
- Vanderah, T. W. *et al.* Mediation of swim-stress antinociception by the opioid *delta*2 receptor in the mouse. *J. Pharmacol. Exp. Therapeut.* **262**, 190–197 (1992).
- Jessell, T. M. & Iversen, L. L. Opiate analgesics inhibit substance P release from rat trigeminal nucleus. *Nature* **268**, 549–551 (1977).
- Siegel, A., Schubert, K. & Shaiku, M. B. Neurochemical mechanisms underlying amygdaloid modulation of aggressive behaviour in the cat. *Aggr. Behav.* **21**, 49–62 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. We thank N. Unwin, W. Wisden, R. Munglani and R. Twyman for reading the manuscript; S. Nakanishi for the NK1 receptor cDNA; J. Ure, A. Jones, C. Roza and L. Singh for technical advice or assistance; S. Ingham for photographic assistance; P. Mantyh for suggestions and discussion. This work was supported by grants from MRC-ROPA, BBSRC, EC and DGICYT.

Correspondence and requests for materials should be addressed to S.P.H. (e-mail: hunt@mrc-lmb.cam.ac.uk).

A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction

Karine Clément^{*†‡}, Christian Vaisse^{*†‡}, Najiba Lahlou[§], Sylvie Cabrol^{||}, Veronique Pelloux^{*}, Dominique Cassuto^{*}, Micheline Gourmelen^{||}, Christian Dina[†], Jean Chambaz[†], Jean-Marc Lacorte[†], Arnaud Basdevant^{*†}, Pierre Bougnères[¶], Yves Leboucq^{||}, Philippe Froguel^{*†} & Bernard Guy-Grand^{*†}

^{*} Laboratoire de Nutrition et Service de Médecine et Nutrition, Hôtel-Dieu place du Parvis Notre Dame, 75004 Paris, France

[†] Institut de Biologie-CNRS EP10, Institut Pasteur de Lille, rue Calmette, 59000 Lille, France

[§] Inserm U342, Hôpital Saint Vincent de Paul et service d'Endocrinologie-Diabète de l'Enfant, avenue Denfert Rochereau, 75014 Paris, France

^{||} Explorations fonctionnelles endocriniennes, Hôpital d'enfant Armand Trousseau, avenue du Dr Arnold Netter, 75012 Paris, France

[¶] CJF INSERM 9508, 15 rue de l'Ecole de Médecine, 75005 Paris, France

[‡] These authors contributed equally to this work.

The adipocyte-specific hormone leptin, the product of the *obese* (*ob*) gene, regulates adipose-tissue mass through hypothalamic effects on satiety and energy expenditure^{1–4}. Leptin acts through the leptin receptor, a single-transmembrane-domain receptor of the cytokine-receptor family^{5–7}. In rodents, homozygous mutations in genes encoding leptin¹ or the leptin receptor⁶ cause early-onset morbid obesity, hyperphagia and reduced energy expenditure. These rodents also show hypercortisolaemia, alterations in glucose homeostasis, dyslipidaemia, and infertility due to hypogonadotropic hypogonadism⁸. In humans, leptin deficiency due to a mutation in the leptin gene is associated with early-onset obesity⁹. Here we describe a homozygous mutation in the human leptin receptor gene that results in a truncated leptin receptor lacking both the transmembrane and the intracellular domains. In addition to their early-onset morbid obesity, patients homozygous for this mutation have no pubertal development and their secretion of growth hormone and thyrotropin is reduced. These results indicate that leptin is an important physiological regulator of several endocrine functions in humans.

We studied a family (HD) with a strong prevalence of morbid obesity occurring early in life. The 19-year-old proband (HD416) is the fourth child of nine siblings. Her parents are second cousins in a consanguineous family of Kabilian origin. She shared her phenotype with two other morbidly obese sisters (HD413 and HD417, who died at age 19) (Fig. 1a). None of the siblings had features of Prader–Willi or Bardet–Biedel syndromes, or had abnormalities in karyotype or in brain computerized-tomography scans. Parents and other siblings do not share the same severe obese phenotype (Fig. 1a). Serum leptin concentrations (Table 1) of the three sisters were much higher than those measured in 842 consecutive subjects consulting in our department ($20 < \text{body mass index (BMI)} < 80 \text{ kg per m}^2$, leptin range $2\text{--}160 \text{ ng ml}^{-1}$). Serum leptin levels of both parents and three of the siblings (HD414, HD418 and HD419) were half those of the affected sisters. Two additional siblings (HD412 and HD415) had leptin levels in the normal range for their BMI.

To determine whether the genes encoding leptin or the leptin receptor (Ob-R) were involved in the extreme obesity of the affected sisters, we studied the genotype of all family members at microsatellite markers at these loci. Four markers located at the *ob* locus showed no segregation of any haplotype with the disease. In contrast, markers (D1S203, D1S220 and D1S198) located at the Ob-R gene locus cosegregated with the morbidly obese phenotype.

We scanned the 18 coding exons of the entire long leptin-receptor isoform for mutations, using single-stranded conformation-dependent polymorphism (SSCP). An abnormal conformer of exon 16 was found in a homozygous state in HD416 and in a heterozygous state in her mother (Fig. 2a). Direct nucleotide sequencing of exon 16 of HD416 identified a G → A base substitution in the splice donor site of exon 16 (Fig. 2b). Both parents and four of the six non-affected siblings were heterozygous for this variant, whereas the other two non-affected siblings were homozygous for the normal allele (Fig. 1b). The lod score for linkage between the G → A mutation and morbid obesity is 3.53, at a recombination fraction of 0, under a recessive model with complete penetrance. No mutation of exon 16 was detected in 402 French Caucasians who were of normal weight or obese¹⁰, confirming that this mutation is not a common one.

We investigated the effect of the Ob-R gene mutation on the encoded messenger RNA. Amplification of a 275-base-pair (bp) complementary DNA region spanning exon 16 (exons 15–17, a

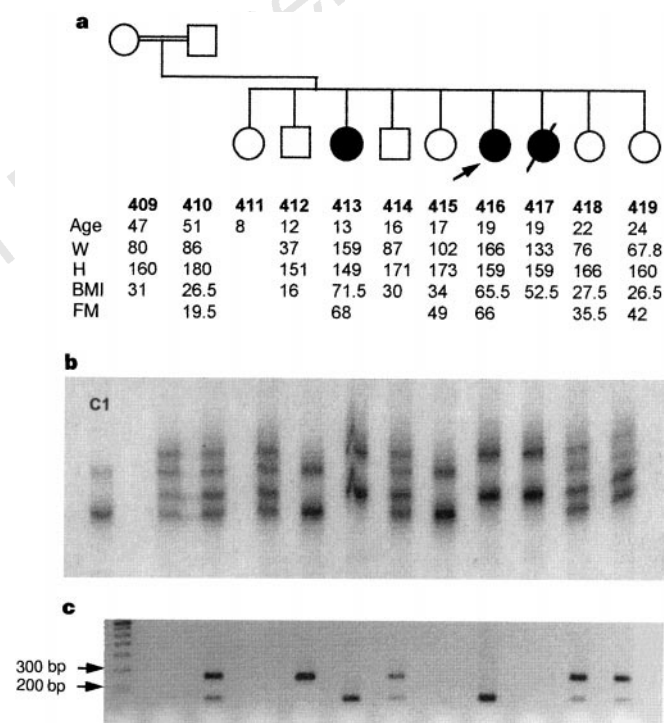


Figure 1 Pedigree and clinical characteristics of the HD family, and analysis of exon 16 of the leptin receptor in this family. **a**, Pedigree and clinical characteristics of the HD family. Filled symbols denote the three affected girls, with an arrow indicating the proband. Identification numbers (HD409, HD410, etc) and available clinical data are shown. Age, years; W, weight in kg; H, height in cm; BMI, kg per m². FM is the percentage of total body fat measured in the father and five siblings using biphonetic absorptiometry (Hologic Device QDR 1000/W, Watham, MA). The father, HD410, is a diabetic who is treated by diet and with an α -glucosidase inhibitor. No clinical data except for age were available for the non-obese youngest sibling, HD411. **b**, SSCP analysis of exon 16 of the human leptin receptor gene in the HD family. SSCP analysis on a 5%, non-denaturing, glycerol-free polyacrylamide gel (electrophoresed at 6–20W for 4–7 h at 4°C) shows a mutation, resulting in a mobility shift, seen in homozygous form in HD413, HD416 and HD417. Of the clinically unaffected siblings, four (HD411, HD414, HD418 and HD419) were heterozygous and two (HD412 and HD415) were homozygous for the wild-type sequence. C1 is a control obese subject who is unrelated to the HD family. **c**, RT-PCR of the region of the human leptin receptor gene spanning exon 16. For clinically unaffected individuals, a PCR product with an expected size of 276 bp was seen. For affected sisters HD416 and HD417 there was a unique band of 170 bp that lacked exon 16. The father and the heterozygous sibling express both mRNA species. mRNA was not available for HD409, HD411, HD415 and HD417.

region common to all transmembrane-domain-containing isoforms) shows that an abnormal Ob-R mRNA is expressed in the heterozygous subjects. This mRNA is not expressed in the normal individuals but is the only form expressed in the affected sisters (Fig. 1c). Direct nucleotide sequencing of the amplified product produced from reverse transcription and polymerase chain reaction (RT-PCR) confirms that the abnormal mRNA results from skipping of exon 16. The receptor mRNA lacking exon 16 potentially encodes a protein of 831 amino acids (Ob-Rhd), which would contain the first 830 amino-terminal amino acids of the normal receptor extracellular domain and extra glutamine at the carboxy terminus. This protein would lack both the transmembrane and the intracellular domains of the Ob-R. The expression of the 824 amino acids of the human Ob-R extracellular domain in COS cells leads to the secretion of a leptin-binding protein that is able to dimerize¹¹. To determine whether Ob-Rhd has similar properties, we assessed how leptin is carried in the serum of the patients by radioimmunoassay quantification of leptin in eluted fractions after gel filtration chromatography. In normal and obese controls, a low proportion (5–20%) of serum leptin circulates in a high-molecular-mass complex, as described^{12,13}. Its apparent relative mass (M_r) is 440K, which is close to that found with similar techniques¹³. In contrast, in the serum of the homozygous and heterozygous mutant patients 80% of leptin circulates as a complex of similar size. Recombinant human leptin displaces ¹²⁵I-labelled leptin from the high-molecular-mass complex, confirming the presence of a specific leptin binding factor. The M_r of the complex could indicate dimerization of Ob-Rhd and also the presence of short circulating forms of Ob-R.

The affected sisters were studied for obesity and short stature during childhood. Their birthweight was normal (3.60, 4.00 and

3.620 kg for HD416, HD413 and HD417, respectively), but severe obesity rapidly developed within the first months of life (Fig. 3). They showed abnormal eating behaviour resembling that seen in Prader–Willi syndrome and in individuals with anatomical damage of the hypothalamic area; behaviour included fighting with other children for food and acting with impulsivity and stubbornness. Repeated psychological evaluations showed emotional lability and social disability but no mental retardation. The resting metabolic rates (RMRs) of HD416 and HD413, evaluated by indirect calorimetry, were 2,510 and 2,181 kcal per 24 h, respectively (the RMR predicted by the Harris and Benedict formula is similar: 2,370 and 2,453 kcal per 24 h in HD416 and HD413, respectively). Their core temperature was normal. Other HD family members are moderately obese but none has a phenotype comparable to that of the affected sisters. HD415, who has the wild-type Ob-R genotype, has the highest BMI of the unaffected siblings (Fig. 1a). The absence of morbid obesity in the heterozygotes indicates that the lack of one normal allele for Ob-R and secretion of an abnormal leptin-binding protein does not affect body-weight regulation by leptin. The leptin levels of the heterozygotes are about half those of the homozygous patients, indicating that the Ob-R mutation might have a codominant effect on the hormone (Table 1). These high leptin concentrations might be related to leptin trapping by the serum leptin-binding factor, resulting in a prolonged half-life of leptin (assuming that the bound leptin is inactive). In homozygotes, the elevated leptin levels might be related to the same mechanism but, in the absence of a functional Ob-R (that is, leptin resistance), an inappropriate oversecretion of leptin is also possible.

The homozygous patients did not spontaneously develop puberty. They had no mammary glands, sparse pubic hair and no

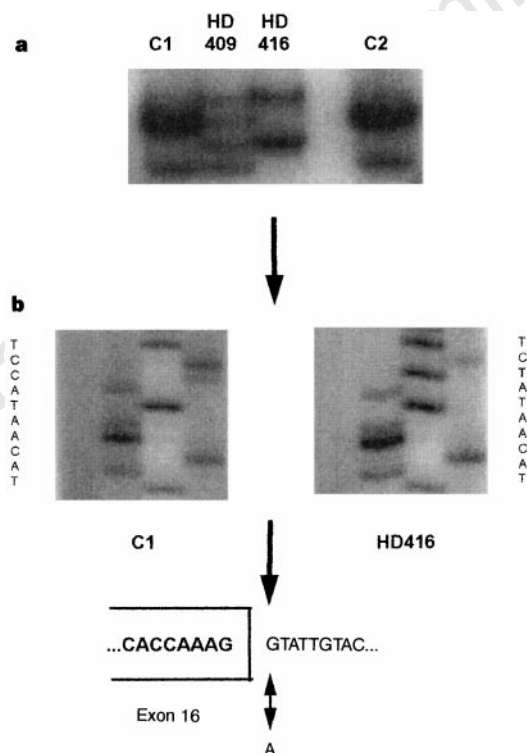


Figure 2 SSCP scanning and DNA-sequence analysis of exon 16 of the human leptin receptor. **a**, An homozygous mobility shift in exon 16 was seen for the proband HD416 as compared with two control obese subjects (C1 and C2) unrelated to the family. Here mother, HD409, was heterozygous for the same variant conformer. **b**, Sequence analysis (reverse complementary strand shown) of the exon 16/intron 16 junction of HD416 and of a control subject. There is a G → A base substitution in the splice donor site of exon 16 in patient HD416.

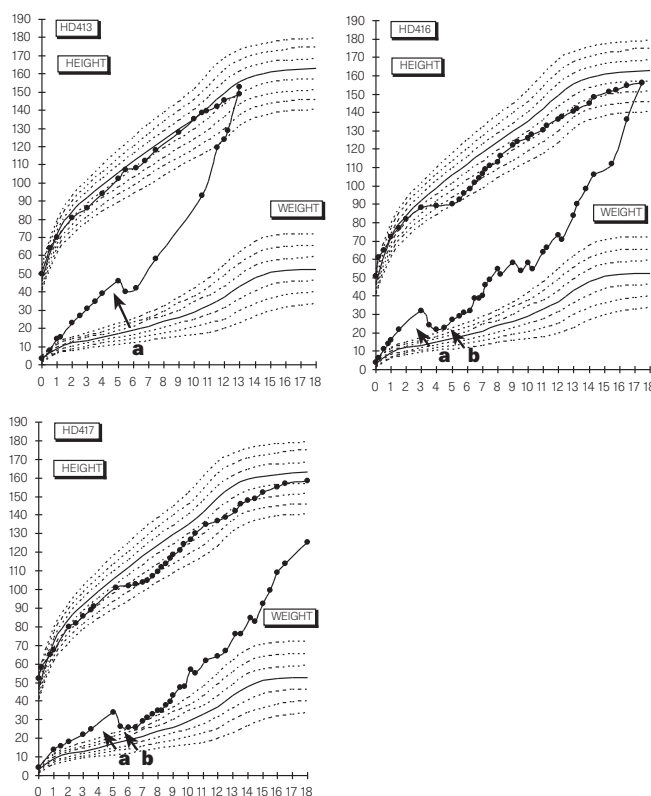


Figure 3 Height and weight curves for the three affected sisters from birth to adult age. The letter **a** indicates a period of food of food-intake restriction. The restrictive diet of 500 kcal per day resulted in weight loss and in a dramatic decrease in growth velocity, which persisted even after food-intake increase and weight regain. The letter **b** indicates the introduction of treatment with levothyroxine and the start of treatment with exogenous growth hormone. The x-axis indicates age in years; the y-axis indicates height in cm of weight in kg.

axillary hair. At age 13.5 (HD413) and 19 (HD416 and HD417), the patients were amenorrhoeic. Basal concentrations of oestradiol were low, and levels of luteinizing hormone and follicle-stimulating hormone (basal and in response to gonadotropin-releasing hormone) remained low or in the normal range (Tables 1 and 2). These features characterize hypogonadism of central origin.

Leptin-deficient *ob/ob* and Ob-R-deficient *db/db* mice are infertile, and leptin is involved in the onset of puberty in rodents^{14–16}. In humans, longitudinal and transversal studies show increasing leptin levels before the onset of puberty^{17,18}, but the leptin-deficient children were too young to confirm an effect of leptin deficiency on gonadotropin secretion⁹. The hypogonadotropic hypogonadism of the sisters with mutated Ob-R shows that signalling through Ob-R is necessary for sexual maturation in women. One functional copy of the human Ob-R gene seems to be enough to allow normal reproductive functions in the heterozygous members of the family. Two heterozygous girls (HD418 and HD419) passed through puberty normally, with onset of menstruation at 13 and 15 years, respectively. The heterozygous males had normal sexual maturation.

The secretion of growth hormone in response to stimulation is generally decreased in obese children¹⁹, and the effects of growth-hormone deficiency are mimicked, except that obese children have normal or slightly increased insulin-like growth factor-I (IGF-I) levels²⁰ and, often, accelerated growth²¹. The affected sisters here showed a mild but significant growth delay during early childhood (Figs 1 and 3), despite normal parental height. No overnight burst of growth hormone was observed and growth-hormone secretion remained low in response to stimulation tests ($<5 \text{ ng ml}^{-1}$) (Table 2). Unexpectedly²⁰, concentrations of IGF-I and IGF-binding protein 3 (IGF-BP3) were decreased (Table 2). Levels of IGF-I and IGF-BP3 increased after administration of growth hormone, as also observed in patients lacking growth hormone (Table 2). This inadequate secretion of growth hormone in the patients with the Ob-R mutation does not occur in leptin-deficient children, who appear to have normal linear growth⁹. In rodents, levels of growth hormone in the pituitary and of growth-hormone-releasing hormone in the hypothalamus are decreased in young *db/db* mice²². Both *ob/ob* and *db/db* mice show stunted growth curves²², and leptin-receptor-deficient obese Zucker rats secrete low levels of growth hormone²³. Administration of leptin antiserum in rodents decreases spontaneous growth-hormone secretion²⁴, indicating a direct effect of leptin on growth.

Low levels of free thyroxine, normal levels of basal thyroid-stimulating hormone (TSH) and a sustained TSH response to

Table 2 Endocrinological data for the three sisters with homozygously mutated Ob-R genes*

	HD416	HD413	HD417	Reference values
GnRH test	(13.5)	(13)	(13.5)	girls Tanner P2
LH (mU ml^{-1}) basal	<0.7	<0.2	<0.8	1.0–4.0
LH peak	1		0.9	6.0–60
FSH (mU ml^{-1}) basal	1	<0.1	1.20	1.0–4.0
FSH peak	4.3		3.2	3.0–20
Oestradiol (pg ml^{-1})	10	17	13	8.0–50
GH peak (ng ml^{-1}) after orthithine and insulin test†	(5)	(4)	(7)	
Normal diet	1.6 & 0.1	0.4 (10.5)	3.5 & 1.4	>10
Overnight Restriction diet	<1	3	1.6	
IGF (ng ml^{-1}) Basal	(5)	(2)	(7)	130 $\pm$ 50 at 2 years 190 $\pm$ 60 at 5 years 210 $\pm$ 60 at 7 years
After GH stimulation	378	182	135	
‡IGF-BP3 Basal (WLB)	(5)			
After GH stimulation	30%			84.5 $\pm$ 1.4%
Thyroid hormones	(5)	(5)	(7)	
FT4 (pg ml^{-1})	7	7	5	15.8 $\pm$ 3.3
TSH ($\mu\text{U ml}^{-1}$)	3	1	5	<5
TRH test	(5)	(5)	(7)	
TSH basal	3	1.4	2	<5
TSH peak	25	28	15	17 $\pm$ 5
TSH 120'	16	13	10	<5
Glucocorticosteroids	(11.5)	(10.5)	(7)	
FLU ($\mu\text{g per 24 h}$)	89 (19)	60	64	<40 prepubertal girls <80 pubertal girls
Cortisol (ng ml^{-1}), 8:00	115 (19)	127 (13.5)		107 $\pm$ 5.2
Cortisol, 16:00	79	78		
ACTH (pg ml^{-1}), 8:00	68	32		6.0–50
ACTH, 16:00	68	26		
OGTT	(3)		(5)	
Glycaemia (mmol l^{-1}) basal	4.8		4.9	3.7–5.9
Peak	5.5		7.4	
120'			4.7	<10
Insulinaemia (ng ml^{-1}) basal	5		<5	<15
Peak	7		49	
120'			8	30–90

* The three sisters were studied several times during infancy at different ages. Ages (in years) at which the sisters were studied are shown in parentheses.

† Normal diet means *ad libitum* diet and restrictive diet means a 700 kcal (25% glucose) diet.

‡ IGF-BP3 in HD416 (measured by WLB, western ligand blot) is given as a percentage of total. IGF-BP3 levels were measured in basal conditions and after three days of administration of growth hormone (GH). LH, luteinizing hormone; FSH, follicle-stimulating hormone; free thyroxine; TRH, thyrotropin-releasing hormone; FLU, free urinary cortisol; ACTH, adrenocorticotrophic hormone; OGTT, oral glucose tolerance test.

Table 1 Recent biological and hormonal characteristics in the HD family

HD	Genotype	Leptin (ng ml^{-1})	Glycaemia (mmol l^{-1})	Triglycerides (mmol l^{-1})	Cholesterol (mmol l^{-1})	Insulin (IV l^{-1})	FSH (IV l^{-1})	LH (IV l^{-1})	Oestradiol (pg ml^{-1})	Testosterone (ng ml^{-1})	IGF-I (ng ml^{-1})
409	MN	362	3.92	0.62	5.54	10	26.6	17.2	7	0.16	164 (221 $\pm$ 84)
410	MN	145	6.66	2.25	2.55	14	1.65	1.51	13	3.95	119 (248 $\pm$ 48)
412	NN	5.6	4.31	0.5	5.0	8	1.25	0.63	<5	0.41	164 (211 $\pm$ 67)
413	MM	670	6.4	1.0	4.26	14	<0.2	<0.1	17	0.54	178 (211 $\pm$ 67)
414	MN	212	4.08	0.75	4.13	16	3.89	1.56	16	3.46	385 (313 $\pm$ 52)
415	NN	88	4.5	0.87	4.26	30	4.11	12.3	216	0.48	357 (418 $\pm$ 93)
416	MM	600	4.53	1.12	4.40	41	5.9	4.00	21	0.19	154 (303 $\pm$ 50)
417	MM	526	6.0	–	–	–	–	–	13	0.22	–
418	MN	240	4.08	0.94	5.80	6	3.44	4.44	100	0.32	203 (303 $\pm$ 50)
419	MN	294	3.75	0.69	4.0	17	5.47	6.52	44	0.50	274 (303 $\pm$ 50)

The family (except HD411) was studied recently. MM and MN are homozygous and heterozygous, respectively, for the Ob-R gene mutation. NN are subjects with the wild-type genotype. IGF-I reference range depends on age, sex and maturation. The other hormonal reference ranges are as follows. Adult men: follicle-stimulating hormone (FSH), 1.0–6.0; luteinizing hormone (LH), 0.8–6.0; oestradiol, 10.0–40.0; testosterone, 3.4–10. Follicular phase (19–40 years): FSH, 2.0–6.0; LH, 1.1–4.5; oestradiol, 19–94; testosterone, 0.10–0.70. Postmenopausal (<60 years): FSH, 25–90; LH, 12–50; oestradiol, 5–50; testosterone, 0.15–0.50. Pubertal boys, stage II: FSH, 0.4–2.8; LH, 0.4–1.9; oestradiol, 6–27; testosterone, 0.20–3.0. Pubertal girls, stage II: FSH, 0.8–4.2; LH, 0.3–2.5; oestradiol, <5 –46; testosterone, <0.05 –0.35.

stimulation with thyrotropin-releasing hormone indicate hypothalamic hypothyroidism in the affected sisters here (Table 2). The leptin-deficient children⁹ present mild elevation of TSH levels and normal levels of free thyroxine, which can be seen in thyroid dysfunction of central origin. Leptin effects on the thyrotrope axis have been shown in normal mice²⁵, but impairment in the thyrotrope axis is not clearly demonstrated in leptin- or Ob-R-deficient rodents.

Morning levels of adrenocorticotrophic hormone and cortisol and tests for dexamethasone and metyrapone were normal in the affected girls, indicating no gross defect in the hypothalamic pituitary adrenal axis. Urine-free cortisol excretions were slightly elevated (Table 2). This is common in obesity²⁶, and may not be caused by the Ob-R mutation.

The sisters homozygous for the Ob-R mutation repeatedly showed normal fasting glycaemia and normal oral glucose tolerance (Table 2) despite their extreme obesity (Fig. 1). Post-glucose-load insulinaemia was normal during childhood. At 13.5 and 19 years, respectively, insulin levels of HD413 and HD416 were similar to those of their moderately obese sister (HD415, with the wild-type genotype) and of French morbidly obese patients¹⁰ (mean insulin levels: 20 ± 12 mIU per ml). Ob-R-deficient patients are similar to leptin-deficient children in showing normal glucose homeostasis. In *ob/ob* and *db/db* mice, the importance of the diabetic phenotype depends on the background. Severe hyperglycaemia is seen on the C57BL/KsJ background^{27,28} and mild diabetes with transient hyperglycaemia and marked hyperinsulinaemia are seen on the C57BL/6J background^{27,28}. All these data indicate that leptin deficiency may not necessarily lead to diabetes. Contrasting with *db/db* and *ob/ob* mice, which are characterized by hypertriglyceridaemia and hypercholesterolaemia, fasting triglyceridaemia and cholesterolaemia are normal in the patients with the Ob-R mutation (Table 1).

We have described a mutation in the human leptin receptor and have analysed the clinical and biological data available for the affected patients. Our results indicate that a functional leptin receptor is required not only for the regulation of body weight but also for sexual maturation and for secretion of growth and thyrotrophic hormones. Leptin is therefore a critical link between energy stores and hypothalamic pituitary functions in humans. □

Methods

Genotyping. Genotypes were determined using microsatellite polymorphisms located at the Ob or at the Ob-R gene loci using standard techniques.

SSCP. Ob-R-coding exons from the HD family and from exon 16 of the controls were amplified in the presence of [α -³²P]-dATP using described primers²⁹ or the following primers: forward exon 1, 5'-TAA ATT TAG AGA CTT ATC TAT AAT CCC-3'; reverse exon 1, 5'-TAA CTA GAA ATA GGA AAT TCT GTT AGC-3'; forward exon 3, 5'-TTT TTT TTA AAT TCA GAT GCA AAC TGG A-3'; reverse 4 I, 5'-ACA TAA GGA GAG TGT CGT; forward exon 6, 5'-AGT AAC GGT TCC ACA TCA ACT TG-3'; forward exon 7I, 5'-CAG AAT GTT TGT CTT CAT CTG ATA TCC-3'; reverse exon 7II, 5'-ATT CAA GTT GTG GAA CAA AAT GAA CA-3'; reverse exon 8, 5'-ATT TTT ATC TTC ACT GTG CCC AC-3'; forward exon 11, 5'-GTA CTT CAG GGC CCC TTT AGA TAC ATA-3'; reverse exon 11, 5'-TTT GAA GAA ATA CTT TTC AGC CAT A-3'. (Coding exons 2, 4, 7/8 and 18 were amplified using 2, 2, 2 and 4 primer pairs, respectively.) The radiolabelled PCR products were electrophoresed overnight in acrylamide gel (with 0% and 10% glycerol) and autoradiographed according to standard SSCP techniques (Fig. 1b).

RT-PCR amplification of human Ob-R cDNA. Total RNA from lymphocytes was prepared using the RNeasy spin kit (Eurogentec, Seraing, Belgium). RNAs were reverse-transcribed using AMV reverse transcriptase (Gibco BRL, Oxford, England). PCR (1.25 mM MgCl₂, annealing temperature 60 °C) was performed for exons 15–17 using the following primers (forward 15, 5'-CAT TTT ATC CCC ATT GAG AAG TA-3'; reverse 17, 5'-CTG AAA ATT AAG TCC TTG TGC CCA G-3').

Sequencing. PCR products were sequenced by AmpliTaq cycle sequencing (Promega, Lyon, France) using [γ -³²P]-end-labelled primers.

Gel-filtration analysis of serum leptin and displacement of bound ¹²⁵I-labelled leptin. Serum (1 ml) was applied to an Ultrogel AcA 44 column (Biosepra, Villeneuve La Garenne, France). Immunoreactive leptin was measured by radioimmunoassay in 0.5-ml fractions. Displacement of ¹²⁵I-labelled leptin was measured after incubation of serum plus 15,000 c.p.m. of ¹²⁵I-labelled leptin, with or without 10 µg of human recombinant leptin (Peninsula, UK).

Determination of molecular mass of leptin-binding factor. Serum (100 µl) was incubated with 15,000 c.p.m. ¹²⁵I-labelled leptin applied to a Sephacryl S300 HR 6/2/98 column, and calibrated with a gel-filtration kit (Pharmacia, St Quentin-Yvelines, France).

Hormonal assays. Hormonal measurements were made using commercially available kits. IGF1 and IGF-BP levels were measured as described³⁰.

Received 30 October 1997; accepted 20 February 1998.

- Zhang, Y. *et al.* Positional cloning of the mouse *obese* gene and its human homologue. *Nature* **372**, 425–432 (1994).
- Halaas, J. L. *et al.* Weight reducing effects of the plasma protein encoded by the *obese* gene (*ob*). *Science* **269**, 543–546 (1995).
- Pelleymounter, M. A. *et al.* Effects of the *obese* gene product on body weight regulation in *ob/ob* mice. *Science* **269**, 540–543 (1995).
- Campfield, L. A., Smith, F. J., Guise, Y., Devos, R. & Burn, P. Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks. *Science* **269**, 546–549 (1995).
- Tartaglia, L. A. *et al.* Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**, 1263–1271 (1995).
- Lee, G. H. *et al.* Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379**, 632–635 (1996).
- Vaisse, C. *et al.* Leptin activation of Stat3 in the hypothalamus of wild-type and *ob/ob* mice but not *db/db* mice. *Nature Genet.* **14**, 95–97 (1996).
- Coleman, D. L. *Obese* and *Diabetes*: two mutant genes causing diabetes-obesity syndromes in mice. *Diabetologia* **14**, 141–148 (1978).
- Montague, C. T. *et al.* Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* **387**, 903–907 (1997).
- Clément, K. *et al.* Association of poorly controlled diabetes with low serum leptin in morbid obesity. *Int. J. Obes.* **21**, 556–561 (1997).
- Liu, C. *et al.* Expression and characterization of a putative high affinity human soluble leptin receptor. *Endocrinology* **138**, 3548–3554 (1997).
- Houck, K. L. *et al.* Evidence for leptin binding to proteins in serum of rodents and humans: modulation with obesity. *Diabetes* **45**, 1638–1643 (1996).
- Diamond, F. B. *et al.* Demonstration of a leptin binding factor in human serum. *Biochem. Biophys. Res. Commun.* **233**, 818–822 (1997).
- Chehab, F. F., Lim, M. E. & Lu, R. Correction of the sterility defect in homozygous obese female mice by treatment with the human recombinant leptin. *Nature Genet.* **12**, 318–320 (1996).
- Chehab, F. F., Mounzih, K., Lu, R. H. & Lim, M. E. Early onset of reproductive function in normal female mice treated with leptin. *Science* **275**, 88–90 (1997).
- Ahima, R. S., Dushay, J., Flier, S. N., Prabakaran, D. & Flier, J. S. Leptin accelerates the onset of puberty in normal female mice. *J. Clin. Invest.* **99**, 391–395 (1997).
- Mantzoros, C. S., Flier, J. S. & Rogol, A. D. A longitudinal assessment of hormonal and physical alterations during normal puberty in boys. Rising leptin levels may signal the onset of puberty. *J. Clin. Endocrinol. Metab.* **82**, 1066–1070 (1997).
- Clayton, P. E. *et al.* Serum leptin through childhood and adolescence. *Clin. Endocrinol.* **46**, 727–733 (1997).
- Bernini, G. P. *et al.* Impaired growth hormone response to insulin-induced hypoglycemia in obese patients: restoration blocked by ritanserin after fenfluramine administration. *Clin. Endocrinol.* **32**, 453–459 (1990).
- Thissen, J. P., Ketelslegers, J.-M. & Underwood, L. Nutritional regulation of the insulin-growth factors. *Endocr. Rev.* **15**, 80–101 (1994).
- Vignolo, M., Naselli, A., Di Battista, E., Mostert, M. & Aicardi, G. Growth and development in simple obesity. *Eur. J. Pediatr.* **147**, 242–244 (1988).
- Bray, G. A. & York, D. A. Hypothalamic and genetic obesity in experimental animals: an autonomic and endocrine hypothesis. *Physiol. Rev.* **59**, 719–790 (1979).
- Tannenbaum, G. S., Lapointe, M., Gurd, W. & Finkelstein, F. A. Mechanism of impaired growth hormone secretion in genetically obese Zucker rats: roles of growth hormone factor and somatostatin. *Endocrinology* **127**, 3080–3095 (1990).
- Carro, E., Senaris, R., Considine, R. V., Casanueva, F. F. & Dieguez, C. Regulation of *in vivo* growth hormone secretion by leptin. *Endocrinology* **138**, 2203–2206 (1997).
- Ahima, R. S. *et al.* Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
- Björntorp, P. Body fat distribution, insulin resistance, and metabolic diseases. *Nutrition* **13**, 795–803 (1997).
- Hummel, K. P., Coleman, D. L. & Lane, P. W. The influence of genetic background on expression of mutations at the diabetes locus in the mouse C57BL/KsJ and C57BL/6J strains. *Biochem. Genet.* **7**, 1–13 (1972).
- Coleman, D. L. & Hummel, K. P. The influence of the genetic background on the expression of the *obese* (*ob*) gene in the mouse. *Diabetologia* **9**, 287–293 (1973).
- Echwald, S. M. *et al.* Amino acid variants in the human leptin receptor—lack of association to juvenile onset obesity. *Biochem. Biophys. Res. Commun.* **233**, 248–252 (1997).
- Hardouin, S. *et al.* Molecular forms of serum insulin-like growth factor (IGF) binding proteins in man: relationships with growth hormone and IGF1s and physiological significance. *J. Clin. Endocrinol. Metab.* **69**, 1291–1301 (1989).

Acknowledgements. This work was supported by the French Ministère de l'Éducation, de la Recherche et de la Technologie, the Direction de la Recherche Clinique/Assistance Publique-Hopitaux de Paris, Programme Hospitalier de Recherche Clinique and the Institut de Recherche Endocrinienne et Métabolique. We thank E. R. Serra, D. Pepin, S. Carrera, P. Boutin, L. Perrin, J. Le Four and Y. Le Bihan for technical help; J. Di Santo for critical reviewing of the manuscript; and the patients and their families for their cooperation. Informed personal and parental consents were obtained and ethical permission was granted by the local ethics committee (CCPRB, Hôtel-Dieu, Paris).

Correspondence and request for materials should be addressed to P.F. (e-mail: froguel@senope.pasteur-lille.fr)

E-cadherin germline mutations in familial gastric cancer

Parry Guilford*, Justin Hopkins*, James Harraway*, Maybelle McLeod†, Ngahiraka McLeod†, Pauline Harawira†, Huriana Taite†, Robin Scoular‡, Andrew Miller§ & Anthony E. Reeve*

* Cancer Genetics Laboratory, Biochemistry Department, University of Otago, PO Box 56, Dunedin, Aotearoa New Zealand

† Kimi Hauora Health Clinic, PO Box 4062, Mt Maunganui South, Aotearoa New Zealand

‡ Tauranga Public Hospital, Tauranga, Aotearoa New Zealand

§ Pathology Department, University of Otago, PO Box 56, Dunedin, Aotearoa New Zealand

The identification of genes predisposing to familial cancer is an essential step towards understanding the molecular events underlying tumorigenesis and is critical for the clinical management of affected families. Despite a declining incidence, gastric cancer remains a major cause of cancer death worldwide¹, and about 10% of cases show familial clustering^{2,3}. The relative contributions of inherited susceptibility and environmental effects to familial gastric cancer are poorly understood because little is known of the genetic events that predispose to gastric cancer. Here we describe the identification of the gene responsible for early-onset, histologically poorly differentiated, high grade, diffuse gastric cancer⁴ in a large kindred from New Zealand (Aotearoa). Genetic linkage analysis demonstrated significant linkage to markers flanking the gene for the calcium-dependent cell–adhesion protein E-cadherin. Sequencing of the E-cadherin gene revealed a G → T nucleotide substitution in the donor splice consensus sequence of exon 7, leading to a truncated gene product. Diminished E-cadherin expression is associated with aggressive, poorly differentiated carcinomas⁵. Underexpression of E-cadherin is a prognostic marker of poor clinical outcome in many tumour types⁶, and restored expression of E-cadherin in tumour models can suppress the invasiveness of epithelial tumour cells^{7,8}. The role of E-cadherin in gastric cancer susceptibility was confirmed by identifying inactivating mutations in other gastric cancer families. In one family, a frameshift mutation was identified in exon 15, and in a second family a premature stop codon interrupted exon 13. These results describe, to our knowledge for the

first time, a molecular basis for familial gastric cancer, and confirm the important role of E-cadherin mutations in cancer.

Familial gastric cancer in a kindred of Maori ethnicity (family A) was originally reported⁹ in 1964 and in the past 30 years over 25 family members have died of this disease (Fig. 1). There is no evidence of an elevated rate of cancer of other organs in this family. The age of death from gastric cancer ranges upwards from 14 years of age, with the majority of cases occurring in people under the age of 40. This is in marked contrast with the general New Zealand population, in which about 80% of gastric carcinomas occur in people older than 60 years. (B. Cox, personal communication).

The pedigree pattern (Fig. 1) is consistent with the dominant inheritance of a susceptibility gene with incomplete penetrance. To identify this putative gene, we carried out a genetic linkage analysis by using microsatellite markers flanking candidate genes for the cancer-susceptibility locus.

One of these candidates was the homophilic cell adhesion molecule E-cadherin. E-cadherin is a transmembrane protein with five tandemly repeated extracellular domains and a cytoplasmic domain that connects to the actin cytoskeleton through a complex with α , β and γ catenins¹⁰. It is important for establishing cell polarity and maintaining normal tissue morphology and cellular differentiation¹⁰.

The linkage analysis found a maximum two-point lod score ($Z_{\max} = 5.04$, $\theta = 0$) with marker *D16S752* (Table 1), which maps within the genetic interval on chromosome 16q22.1 containing the E-cadherin gene¹¹. Genotyping of five other markers¹² in the vicinity of E-cadherin identified additional significantly linked markers (Table 1). A conserved haplotype spanning 9 centimorgans¹² from *D16S3019* to *D16S3138* was consistently inherited with the disease. This haplotype was also present in all obligate carriers of the

Table 1 Linkage of the gastric cancer susceptibility gene to markers flanking E-cadherin

Marker	Lod scores		Recombination fraction (θ)
	Equal allele frequencies	Kindred allele frequencies	
<i>D16S752</i>	5.04	4.04	0
<i>D16S3043</i>	2.01	2.34	0.05
<i>D16S3019</i>	2.28	1.57	0
<i>D16S3095</i>	4.90	4.07	0
<i>D16S3083</i>	2.79	2.16	0
<i>D16S3138</i>	3.32	2.68	0

Two-point lod scores were calculated assuming either equal allele frequencies or, in a conservative approach (in the absence of allele frequencies for the study population), using the actual frequencies observed in the study kindred.

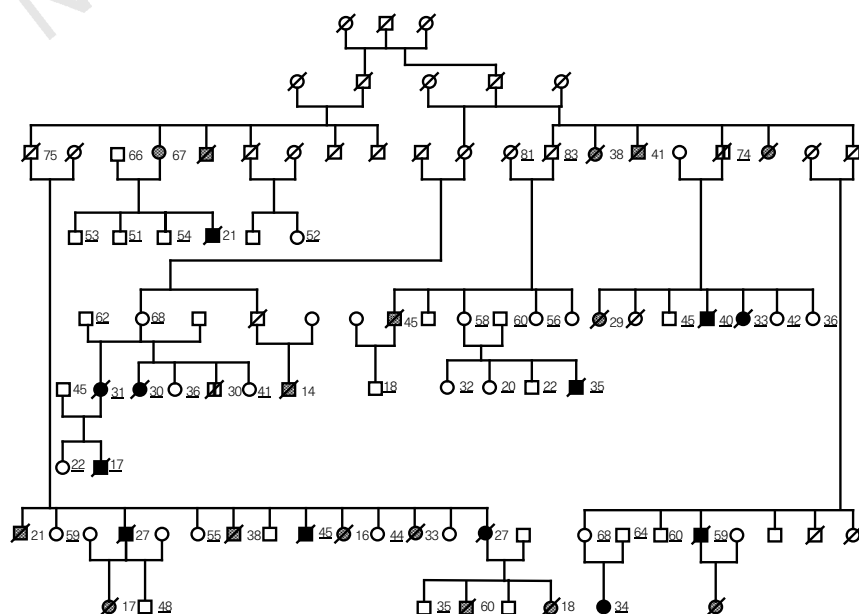


Figure 1 Gastric cancer kindred (family A). Where known, the individual's age is indicated to the right of the symbols. The age is underlined if a blood or biopsy sample was available. General symbols: squares, males; circles, females; all symbols with a diagonal, deceased. Solid symbols: gastric carcinoma, pathology available; dotted symbols: gastric carcinoma, pathology unavailable; vertical stripes: colorectal cancer. Some details unrelated to the mode of inheritance of the disease have been modified to protect the privacy of individual family members.

susceptibility gene, in two individuals affected by colorectal cancer (aged 30 and 74 years) and in a proportion of the unaffected individuals. The proportion of individuals with the haplotype who were affected by the age of 60 provided an approximation of 70% for the penetrance of the susceptibility gene in this kindred.

Mutation analysis of E-cadherin exons 2–16 using the single-stranded conformational polymorphism (SSCP) technique¹³ revealed a band shift in exon 7 (Fig. 2a) in DNA extracted for lymphocytes of two affected people and four obligate carriers of the susceptibility gene. Direct sequencing of exon 7 identified a G → T transversion at the last nucleotide (position 1,008) of this exon (Fig. 2b). The SSCP band shift was not observed in 150 unrelated chromosomes.

This mutation may have two consequences for E-cadherin expression and function. First, the mutated base forms part of the splice donor site^{14,15} for exon 7. To examine the possible effect of the G → T transversion on splicing, we used exon linking RT-PCR (polymerase chain reaction with reverse transcription) on stomach biopsy material taken from an affected family member. In addition to the expected wild-type product of 180 base pairs (bp) derived from exons 7 and 8, a minor band of 187 bp was observed (data not shown). Nucleotide sequencing revealed that all clones derived from the 187-bp product contained the mutation 1008T and a 7-bp insertion derived from intronic sequence between the normal splice donor site and an adjacent cryptic splice site (Fig. 2c). This insertion is predicted to generate a premature stop codon in exon 8 of E-cadherin. The absence of the wild-type nucleotide at position 1,008 associated with the 7-bp insertion argues that use of the cryptic

splice site is limited to the mutant transcript. Cryptic splicing of the 1008T transcript occurs with a high efficiency: nucleotide sequencing revealed that only 1/20 clones derived from the normal 180-bp PCR product contained the 1008T mutation. The presence of a somatic mutation at position 1,008 of E-cadherin in human cancer has been reported previously: A G → A mutation at position 1,008 has been found in a histologically diffuse gastric carcinoma, leading to cryptic splice-site activation and transcript instability¹⁶.

The second consequence of the G → T transversion is the substitution of a glutamic acid residue at position 336 with aspartic acid in the residual correctly spliced transcript. Glu 336 is located in one of the LDRE motifs (single-letter amino-acid code) which are part of E-cadherin's four calcium-binding pockets¹⁷. Calcium binding is required for dimerization and rigidification of E-cadherin and provides protection from proteolytic degradation¹⁸. The LDRE motif is highly conserved¹⁷, not only among vertebrates but also in *Drosophila*¹⁹, suggesting that a Glu → Asp mutation at this position is important for the correct functioning of the protein.

To confirm the role of E-cadherin in inherited gastric cancer susceptibility, we searched for germline mutations in this gene in two other families (families B and C) with early-onset, histologically

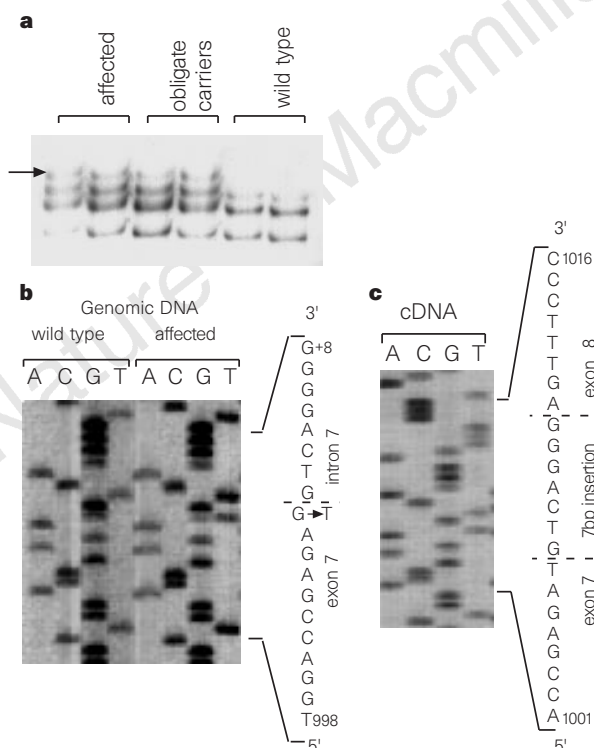


Figure 2 Identification of the mutation in family A. **a**, SSCP pattern of exon 7 in E-cadherin gene. The SSCP band pattern of two affected people, two obligate carriers and two unaffected spouses (wild type) are shown. The additional band in the affected and obligate carrier samples is indicated by the arrow. **b**, Genomic DNA sequence analysis of the exon-intron boundary of exon 7 showing the wild-type sequence and the sequence from an affected person heterozygous for the G → T transversion. The position of the exon-intron boundary is marked. **c**, Example of the sequence analysis of the 187-bp RT-PCR cDNA product showing the E-cadherin exon 7-8 boundary. The nucleotide 1,008 mutation and the 7-bp insertion of intronic DNA are marked.

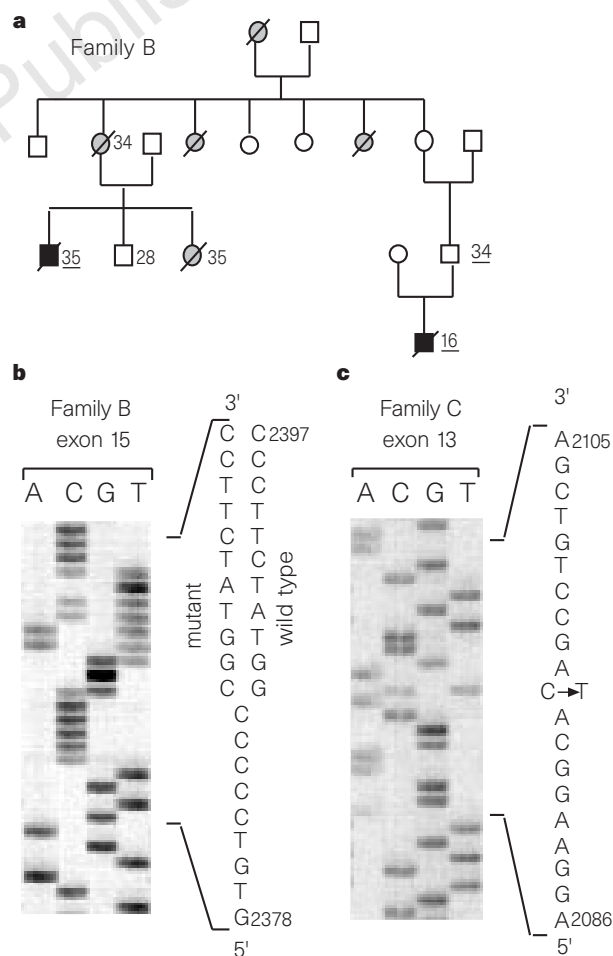


Figure 3 Mutations in families B and C. **a**, Abbreviated pedigree of family B. The symbols are as for Fig. 1. Grey shaded symbols: unconfirmed gastric carcinoma. **b**, Exon 15 DNA sequence (family B) showing the insertion of an additional C in the run of cytosines between positions 2,382-2,386. **c**, Exon 13 sequence (family C) showing the C → T transition at nucleotide 2,095.

Table 2 E-cadherin germline mutations and polymorphisms in gastric cancer families

Family	Nucleotide position (exon)	Mutation	Type
A	1,008 (7)	G → T	Splice site
B	2,382–2,386 (15)	C insertion	Frameshift
C	2,095 (13)	C → T	Premature termination (TAG)
B	1,409 (10)*	C → T	Codon 470: Thr → Ile
A, C	intron 12†	C → T	Silent
A, B, C	2,076 (13)‡	C → T	Silent

* This mutation did not segregate with the disease in family B.

† Located 13 nucleotides upstream of the exon.

‡ This polymorphism has been reported previously (see, for example, refs 13, 24). Nucleotide positions are as described in ref. 17.

diffuse gastric cancer. SSCP analysis of exons 2–16 amplified from lymphocyte DNA was carried out on two affected individuals and one obligate carrier from family B (Fig. 3a) and the proband of family C. A band shift was observed in exon 15 in the three members of family B who were tested. Direct sequencing of exon 15 showed that all three individuals were heterozygous for the insertion of an additional C residue in a run of five cytosines at positions 2,382–2,386 (Fig. 3b). The resulting frameshift leads to an E-cadherin molecule lacking about half of its cytoplasmic domain.

The proband of family C (aged 30 years) showed an SSCP band shift in exon 13. Direct sequencing identified a heterozygous C → T transition at nucleotide 2,095 which converted Gln 699 to a TAG stop codon (Fig. 3c). This inactivating mutation would result in an E-cadherin peptide lacking both the transmembrane and cytoplasmic domains. In addition to the inactivating mutations in families A, B and C, we found silent mutations and one missense mutation which did not segregate with the phenotype (Table 2).

E-cadherin germline mutations have not been previously observed in familial cancer. However, somatic mutations have been identified in sporadic histologically diffuse gastric carcinomas^{20–22}, lobular breast cancers^{13,23} and carcinomas of the endometrium and ovary²⁴. Although a lack of sufficient pure tumour material, due to the high level of contamination with normal cells, precluded our search for loss-of-heterozygosity (LOH), inactivation of both E-cadherin alleles by mutation and LOH^{13,20,22,23} is common in sporadic tumours. Epigenetic inactivation of the E-cadherin gene by CpG hypermethylation has also been observed in several tumour types^{25,26}.

The process of tumorigenesis is driven by the progressive accumulation of somatic mutations in a number of genes. Our identification of a pre-existing mutation in a gene widely regarded as an invasion-suppressor gene⁶ emphasizes that the order of accumulation of mutations leading to tumour formation may belie the observed histopathological progression. The early accumulation of mutations associated with late events of tumorigenesis (such as invasion and metastasis) may explain the unpredictable clinical progression of certain tumour types.

Loss of E-cadherin function and the concomitant disruption of normal cell–cell adhesion could also play a role in the initiation of cell proliferation by allowing escape from growth-control signals. Alternatively, the cytoplasmic domain of E-cadherin may modulate the Wnt signalling pathway by inhibiting the availability of free cytoplasmic β-catenin, in a manner complementary to the product of the colorectal tumour-suppressor gene APC²⁷.

Regardless of the mechanism, our demonstration of germline E-cadherin mutations (Table 2) in families predisposed to diffuse gastric cancer demonstrates a direct role for an intercellular adhesion protein in cancer susceptibility. The high frequency of inactivation of the E-cadherin gene in many types of sporadic tumours^{5,6} suggests that this gene may also confer inherited susceptibility to other cancers. □

Methods

Genotyping. DNA extracted from blood and biopsy samples²⁸ was genotyped

using standard conditions¹² in reactions containing 0.2 U AmpliTaq Gold (Perkin Elmer) and 25 pmol infrared labelled (IR800) forward primer (MWG Biotech). Products were analysed on a LiCor 4000L DNA sequencer.

SSCP analysis. SSCP mutation analysis was carried out as described¹³. PCR products were electrophoresed at room temperature through a 6% non-denaturing polyacrylamide gel without added glycerol; products were detected by autoradiography.

RT-PCR. Total RNA was extracted²⁹ from frozen biopsy material and reverse-transcribed using SuperScript II (Gibco BRL) according to the manufacturer's instructions. Nucleotide position 1,008 was PCR-amplified from the cDNA using a forward primer within exon 7 (5'-taa cag gaa cac agg agt cat ca-3') and a reverse primer from exon 8 (5'-gtg gtg gga ttg aag atc gg-3'). Reactions contained 4 mM MgCl₂ and 0.2 U AmpliTaq Gold and were cycled as follows: (95 °C for 10 min) 1 cycle, and (95 °C for 15 s, 57 °C for 45 s, and 72 °C for 10 s) for 35 cycles.

Plasmid and direct sequencing. RT-PCR products were eluted from a 6% polyacrylamide denaturing gel, re-amplified with the original primers using Pwo polymerase (Boehringer-Mannheim) and ligated into the EcoRV site of Bluescript. Template for direct sequencing of mutations was produced from lymphocyte genomic DNA by PCR using the SSCP antisense primers and the sense primers¹³ with an added 5' leader corresponding to the T3 sequencing primer. Plasmid and direct sequencing were carried out using Thermosequase (Amersham) and an IR800 labelled (MWG Biotech) T3 primer (3 pmol per reaction). Products were analysed on a LiCor 400L DNA sequencer.

Linkage analysis. Two-point lod scores were calculated using MLINK of the LINKAGE 5.1 package³⁰. A gene frequency of 10⁻⁴ was assumed for the disease gene. Age-dependent penetrance was taken into account; seven liability classes were obtained from the cumulative age of onset curve: 0.18 for individuals from 0–20 years, 0.24 (21–25 years), 0.34 (26–30 years), 0.48 (31–35 years), 0.56 (36–40 years), 0.64 (41–45 years) and 0.70 (>46 years). Variation of the maximum penetrance from 60–80% did not change the significance of the results.

Received 12 December 1997; accepted 23 January 1998.

- Howson, C. P., Hiyama, T. & Wynder, E. L. The decline in gastric cancer: Epidemiology of an unplanned triumph. *Epidemiol. Rev.* **8**, 1–27 (1986).
- La Vecchia, C., Negri, E., Franceschi, S. & Gentile, A. Family history and the risk of stomach and colorectal cancer. *Cancer* **70**, 50–55 (1992).
- Zangheri, G. et al. Familial occurrence of gastric cancer in the 2-year experience of a population-based registry. *Cancer* **66**, 2047–2051 (1990).
- Lauren, P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol. Microbiol. Scand.* **64**, 31–49 (1965).
- Shiozaki, H., Oka, H., Inoue, M., Tamura, S. & Monden, M. E-cadherin mediated adhesion system in cancer cells. *Cancer* **77**, 1605–1613 (1995).
- Bracke, M. E., Roy, F. M. & Mareel, M. M. The E-cadherin/catenin complex in invasion and metastasis. *Curr. Topics Microbiol. Imm.* **213**, 123–161 (1996).
- Frixen, E. H. et al. E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. *J. Cell Biol.* **113**, 173–185 (1991).
- Vlemminkx, K., Vakaet, L., Mareel, M., Fiers, W. & Roy, F. V. Genetic manipulation of E-cadherin expression by epithelial tumour cells reveals an invasion suppressor role. *Cell* **66**, 107–119 (1991).
- Jones, E. G. Familial gastric cancer. *NZ Med. J.* **63**, 287–296 (1964).
- Grunwald, G. B. The structural and functional analysis of cadherin calcium-dependent cell adhesion molecules. *Curr. Opin. Cell Biol.* **5**, 797–805 (1993).
- GDB (TM) Human Genome Database. (John Hopkins University, Baltimore, Maryland).
- Dib, C. et al. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
- Berx, G. et al. E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J.* **14**, 6107–6115 (1995).
- Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **55**, 1119–1150 (1986).
- Weil, D. et al. The autosomal recessive isolated deafness, DFNB2, and the Usher 1B syndrome are allelic defects of the myosin-VIIA gene. *Nature Genet.* **16**, 191–193 (1997).
- Oda, T. et al. E-cadherin gene mutations in human gastric carcinoma cell lines. *Proc. Natl Acad. Sci. USA* **91**, 1858–1862 (1994).
- Berx, G. et al. Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1). *Genomics* **26**, 281–289 (1995).
- Nagar, B., Overduin, M., Ikura, M. & Rini, J. M. Structural basis of calcium-induced E-cadherin rigidification and internalization. *Nature* **380**, 360–364 (1996).
- Mahoney, P. A. et al. The fat tumour suppressor gene in *Drosophila* encodes a novel member of the cadherin gene superfamily. *Cell* **67**, 853–868 (1991).
- Becker, K.-F. et al. E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Res.* **54**, 3845–3852 (1994).
- Muta, H. et al. E-cadherin mutations in signet ring cell carcinoma of the stomach. *Jap. J. Cancer Res.* **87**, 843–848 (1996).
- Tamura, G. et al. Inactivation of the E-cadherin gene in primary gastric carcinomas and gastric carcinoma cell lines. *Jap. J. Cancer Res.* **87**, 1153–1159 (1996).
- Berx, G. et al. E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. *Oncogene* **13**, 1919–1925 (1996).
- Risinger, J. I., Berchuck, A., Kohler, M. F. & Boyd, J. Mutations of the E-cadherin gene in human gynaecologic cancers. *Nature Genet.* **7**, 98–102 (1994).

25. Yoshiura, K. *et al.* Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. *Proc. Natl Acad. Sci. USA* **92**, 7416–7419 (1995).
26. Graff, J. R. *et al.* E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. *Cancer Res.* **55**, 5195–5199 (1995).
27. Morin, P. J. *et al.* Activation of β catenin-Tcf signalling in colon cancer mutations in β catenin or APC. *Science* **275**, 1787–1792 (1997).
28. Banerjee, S. K., Makdisi, W. F., Weston, A. P., Mitchell, S. M. & Campbell, D. R. Microwave-based DNA extraction from paraffin-embedded tissue for PCR amplification. *Biotechniques* **18**, 768–773 (1995).
29. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analyt. Biochem.* **162**, 156–159 (1987).
30. Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. Multilocus linkage analysis in humans: Detection of linkage and estimation of recombination. *Am. J. Hum. Genet.* **37**, 482–498 (1985).

Acknowledgements. We acknowledge the support of the New Zealand Lottery Grants Board, Health Research Council of New Zealand, Maurice and Phyllis Paykel Trust, Healthcare Otago Charitable Trust and the Cancer Society of New Zealand; we thank G. Barbezat, J. Cutfield, R. J. M. Gardner, D. Perez, T. Sutton, E. Richardson, D. Shaw, L. Scanlon, J. Ratema, J. Dunphy, I. Morison, M. Eccles for advice and the participating families.

Correspondence and requests for materials should be addressed to P.G. (e-mail: parry.guilford@stonebow.otago.ac.nz).

Stabilization of wild-type p53 by hypoxia-inducible factor 1 α

Won G. An*, Meera Kanekal*, M. Celeste Simon†, Emin Maltepe‡, Mikhail V. Blagosklonny* & Leonard M. Neckers§

* Department of Cell and Cancer Biology, Medicine Branch, NCI, NIH, Bethesda, Maryland 20892, USA

† Departments of Medicine and Molecular Genetics and Cell Biology, and

‡ Department of Pathology, Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637, USA

§ Medicine Branch, NCI, NIH, Key West Facility, 9610 Medical Center Drive, Room 300, Rockville, Maryland 20850, USA

Although hypoxia (lack of oxygen in body tissues) is perhaps the most physiological inducer of the wild-type p53 gene¹, the mechanism of this induction is unknown. Cells may detect low oxygen levels through a haem-containing sensor protein². The hypoxic state can be mimicked by using cobalt chloride and the iron chelator desferrioxamine^{2–5}: like hypoxia, cobalt chloride and desferrioxamine activate hypoxia-inducible factor 1 α (HIF-1 α) (ref. 6), which stimulates the transcription of several genes that

are associated with hypoxia^{6–9}. Here we show that these treatments induce accumulation of wild-type p53 through HIF-1 α -dependent stabilization of p53 protein. Induction of p53 does not occur in either a mutant hepatoma cell line that is unable to induce HIF-1 α (ref. 10) or embryonic stem cells derived from

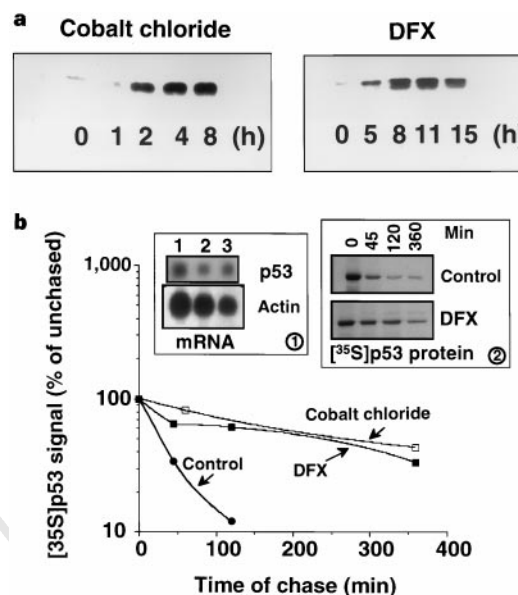
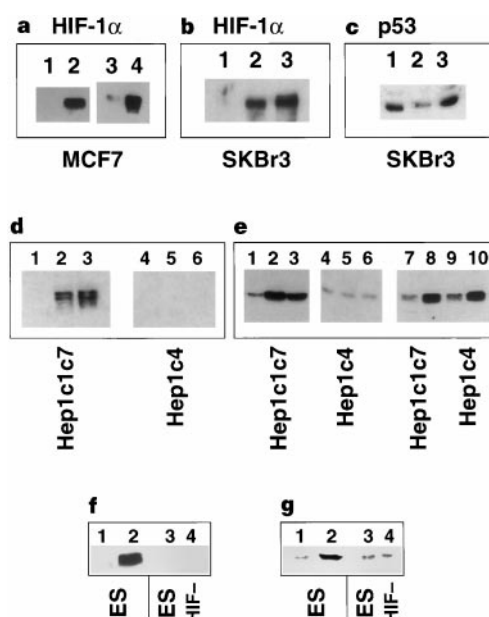


Figure 1 Cobalt chloride and desferrioxamine cause p53 protein stabilization. **a**, Cobalt chloride and desferrioxamine induce wild-type p53 protein accumulation in MCF7 cells. Wild-type p53 was measured by western blot at various times after addition of cobalt chloride or desferrioxamine (DFX). **b**, Cobalt chloride and desferrioxamine increase wild-type p53 protein levels by stabilization. Cobalt chloride or desferrioxamine was added to MCF7 cells for 6 h, the half-life of p53 was then determined by [³⁵S]-methionine pulse-chase analysis. p53-band intensity was quantified by image analysis. Band intensities are expressed as a percentage of the control signal (pulse only, no chase). Inset 1 depicts a northern blot showing p53 mRNA levels (upper panel) in untreated (lane 1), cobalt-chloride-treated (lane 2), or desferrioxamine-treated (lane 3) MCF7 cells 6 h after treatment. The blot was stripped and reprobed for actin (lower panel). Inset 2 shows a representative result of [³⁵S]-methionine pulse-chase analysis in untreated and desferrioxamine-treated cells.

Figure 2 Induction of p53 by cobalt chloride and desferrioxamine requires concomitant induction of HIF-1 α . **a–c**, Cobalt chloride and desferrioxamine induce HIF-1 α protein accumulation in MCF7 and SKBr3 cells, but do not induce accumulation of mutated p53 in SKBr3 cells. Cells were exposed to either cobalt chloride or desferrioxamine for 6 h and lysed; western blots were probed for either HIF-1 α (MCF7 and SKBr3 cells) or mutated p53 (SKBr3 cells). Untreated cells are shown in lanes **a1**, **a3**, **b1** and **c1**; cobalt-chloride-treated cells are shown in lanes **a2** of **a–c**; desferrioxamine-treated cells are shown in lanes **a4**, **b3** and **c3**. **d**, Cobalt chloride and desferrioxamine induce HIF-1 α accumulation in Hepa1c17 but not in Hepa1c4 cells. Cells were treated with cobalt chloride or desferrioxamine for 6 h and nuclear extracts were western blotted to locate HIF-1 α . Untreated cells are shown in lanes 1 and 4; cobalt-chloride-treated cells are shown in lanes 2 and 5; and desferrioxamine-treated cells are shown in lanes 3 and 6. **e**, Cobalt chloride and desferrioxamine induce p53 accumulation in Hepa1c17 but not Hepa1c4 cells, whereas proteasome inhibition induces p53 accumulation in both cell lines. Cells were treated with either cobalt chloride or desferrioxamine and blotted for p53 as in Fig. 1. Untreated cells are shown in lanes 1 and 4; cobalt-chloride-treated cells are shown in lanes 2 and 5; and desferrioxamine-treated cells are shown in lanes 3 and 6. Alternatively, both cell lines were exposed to the proteasome inhibitor ALLN for 6 h and blotted for p53. Untreated cells are shown in lanes 7 and 9, and ALLN-treated cells are shown in lanes 8 and 10. **f**, Desferrioxamine induces HIF-1 α protein accumulation in normal ES cells, but not in ES HIF[−] cells. Lanes 1 and 3 show untreated cells; lanes 2 and 4 show results from cells exposed to desferrioxamine for 6 h. **g**, Desferrioxamine induces p53 accumulation in normal ES cells, but not in ES HIF[−] cells. Lanes 1 and 3 show untreated cells; lanes 2 and 4 show cells exposed to desferrioxamine for 6 h.



mice lacking HIF-1 β (ref. 11). HIF-1 α is found in p53 immunoprecipitates from MCF7 cells that express wild-type p53 and are either hypoxic or have been exposed to desferrioxamine. Similarly, anti-haemagglutinin immunoprecipitates from lysates of normoxic PC3M cells that had been co-transfected with haemagglutinin-tagged HIF-1 α and wild-type p53 also contain p53. Transfection of normoxic MCF7 cells with HIF-1 α stimulates a co-transfected p53-dependent reporter plasmid and increases the amount of endogenous p53. Our results suggest that hypoxic induction of transcriptionally active wild-type p53 is achieved as a result of the stabilization of p53 by its association with HIF-1 α .

MCF7 cells were exposed to either cobalt chloride or desferrioxamine for 1–15 h and levels of p53 protein were determined by western blotting (Fig. 1a). Both agents caused p53 accumulation, which peaked between 4 and 8 h after treatment. Expression of p53 protein induced by both cobalt and desferrioxamine gave p53 with a prolonged half-life (Fig. 1b, graph and inset 2). In untreated MCF7 cells, the half-life of p53 was ~25 min, whereas the protein that accumulated after 6 h exposure to either cobalt or desferrioxamine showed a half-life of >120 min. Similar results were obtained with 1A9 cells, which also express wild-type p53 (data not shown). Northern blot analysis of MCF7 cells exposed to either cobalt or desferrioxamine for 6 h showed no increase in the amount of p53 messenger RNA (Fig. 1b, inset 1), indicating that these hypoxia-mimicking agents caused accumulation of p53 protein through protein stabilization.

Treatment of MCF7 cells 6 h with cobalt or desferrioxamine, or for making them hypoxic, also led to a marked accumulation of HIF-1 α protein (Figs 2a, 3a), as seen in many other cell types^{7,12}. Similar induction of HIF-1 α was seen in SKBr3 cells (Fig. 2b), which express mutated p53. In contrast to the results obtained with MCF7 cells containing wild-type p53, mutated p53 in SKBr3 cells was not

induced to accumulate by either cobalt or desferrioxamine (Fig. 2c).

The hepatoma cell lines Hepa1c1c7 and Hepa1c4 are frequently used to assess the role of HIF-1 α in the mediation of hypoxic responses; Hepa1c4 cells are variants of Hepa1c1c7 cells that do not induce HIF-1 α accumulation upon hypoxic stimulation^{3,10,13–16}. We tested the response of p53 to cobalt and desferrioxamine in both Hepa1c1c7 and Hepa1c4 cells. As expected, both hypoxic stimuli induced accumulation of HIF-1 α in Hepa1c1c7 cells, but neither was able to do so in Hepa1c4 cells (Fig. 2d). Both cobalt and desferrioxamine (6 h exposure) induced p53 protein accumulation in Hepa1c1c7 cells (Fig. 2e; compare lanes 1–3), but neither stimulus was able to do so in Hepa1c4 cells (Fig. 2e, lanes 4–6). As wild-type p53 is degraded by the proteasome, proteasome inhibition results in rapid accumulation of the wild-type protein^{17,18}. As shown in Fig. 2e, p53 markedly accumulated within 6 h of proteasome inhibition in both cell lines (compare lanes 7–10), thus demonstrating a normal response of p53 levels to proteasome inhibition in cells lacking HIF-1 α . To confirm the functional activity of wild-type p53 in Hepa1c4 cells, we transfected these cells with the reporter plasmid PG13-luciferase, which is responsive to wild-type p53. When normalized to the activity seen following transfection with a pGL2-control plasmid (a luciferase reporter that is driven by simian virus 40), PG13 activity in Hepa1c4 cells was equivalent to that seen in Hepa1c1c7 cells (and to that seen in MCF7 cells containing wild-type p53), and both cell lines showed >2 logs more reporter activity specific to wild-type p53 than that found in SKBr3 cells containing mutated p53 (Table 1).

A targeted disruption of the *Arnt* (*HIF-1 β*) locus in the mouse has been generated¹¹. *HIF-1 β* ^{-/-} embryonic stem (ES) cells do not induce HIF transcriptional activity in response to hypoxia, nor are they capable of upregulating expression of a variety

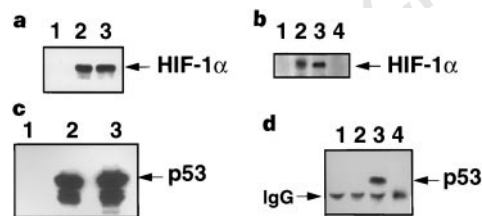


Figure 3 p53 and HIF-1 α proteins associate directly with each other. **a, b**, HIF-1 α protein co-precipitates with wild-type p53 from desferrioxamine (DFX)-treated, or hypoxic, MCF7 cells. Cells were treated with desferrioxamine, or exposed to hypoxic conditions, for 6 h; p53 was then immunoprecipitated from nuclear extracts. HIF-1 α protein levels were determined by western blot in both total extract (**a**) and p53 immunoprecipitates (**b**). In **a**, untreated cells are depicted in lane 1, desferrioxamine-treated cells in lane 2, and hypoxia-exposed cells in lane 3. In **b**, lanes 1 and 2 represent immunoprecipitations from desferrioxamine-treated cells; in lane 1, an irrelevant IgG2 α -isotype control antibody was used as the immunoprecipitating antibody, and in lane 2 antibody against p53 was used. In lane 3, p53 immunoprecipitates were produced from cells exposed to hypoxia. Lane 4 depicts p53 immunoprecipitates from untreated MCF7 cells. **c, d**, Wild-type p53 co-precipitates with HA-HIF-1 α in transiently transfected PC3M cells. PC3M cells were transfected with wild-type p53, alone or with HA-HIF-1 α . Nuclear extracts were prepared 24 h later and immunoprecipitated with anti-HA antibody. Levels of p53 protein were monitored by western blotting in either total extract (**c**) or anti-HA immunoprecipitates (**d**). Untransfected cells are shown in lane 1; cells transfected with wild-type p53 only are shown in lane 2; and cells transfected with both wild-type p53 and HA-HIF-1 α are shown in lane 3. Lane 4 of **d** shows immunoprecipitation using anti-HA antibody of cells transfected with HA-HIF-1 α only.

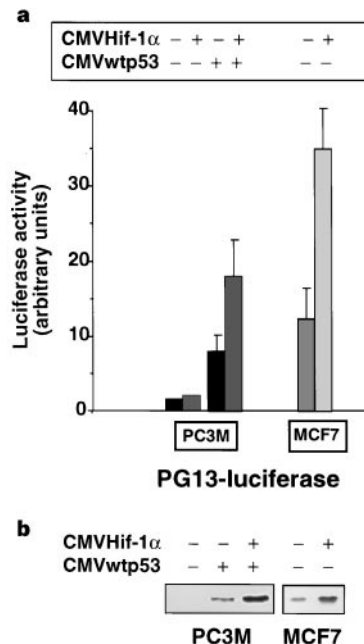


Figure 4 HIF-1 α potentiates PG13-luciferase activity that depends on wild-type p53 (wtp53). Co-transfection of HIF-1 α and wtp53 (through use of expression plasmids CMVHif-1 α and CMVwtp53) augments p53-dependent PG13-luciferase activity in p53-null PC3M cells (**a**), while also increasing the steady-state level of transfected p53 protein (**b**). Transfection of HIF-1 α alone potentiates PG13-luciferase activity in wtp53-expressing MCF7 cells (**a**), and increases the level of endogenous p53 protein (**b**). Luciferase activity and steady-state levels of p53 were determined in cell lysates 16 h after transfection. Luciferase activity is expressed as the mean and standard deviation determined from three experiments. **b**, A representative western blot for p53 (using one-fifth of the protein used in the experiment shown in Fig. 3c).

Table 1 Wild-type p53 functional activity in a variety of cell lines

Cell line	PG13-luciferase activity*
ES	1.3
ES HIF ⁻	0.76
Hepa1c1c7	0.69
Hepa1c4	1.03
MCF7	2.05
SKBr3	0.003

* PG13-luciferase activity was normalized to activity obtained with PGL2-control plasmid.

of hypoxia-responsive genes¹¹. We observed that desferrioxamine is unable to induce accumulation of HIF-1 α protein in HIF-1 β ^{-/-} ES (ES HIF⁻) cells. Under identical conditions, desferrioxamine induced accumulation of HIF-1 α protein in normal (HIF-1 β ^{+/+}) ES cells (Fig. 2f). Desferrioxamine also stimulated p53 accumulation in normal ES cells, but not in ES HIF⁻ cells (Fig. 2g). To assess the functional activity of wild-type p53 in these cells, we transfected them with either PG13-luciferase or pGL2-control plasmids, as described above. When normalized to pGL2-control light output, both normal ES cells and ES HIF⁻ cells displayed levels of PG13 activity that were similar to the levels found in MCF7, Hepa1c1c7 and Hepa1c4 cells (Table 1).

As p53 accumulation seems to be a result of protein stabilization, and therefore not dependent on HIF-1 α -mediated transcriptional activity, we tested whether a direct association of HIF-1 α and p53 proteins could explain p53 accumulation. After 6 h of either desferrioxamine treatment or exposure to hypoxia, we immunoprecipitated p53 from MCF7 cells expressing wild-type p53, subjected the immunoprecipitates to SDS-polyacrylamide gel electrophoresis, and probed the nitrocellulose membrane with an antibody to HIF-1 α (Fig. 3a, b). HIF-1 α was detectable in p53 immunoprecipitates obtained from either desferrioxamine-treated or hypoxia-exposed cells (Fig. 3b, lanes 2, 3). Immunoprecipitation of p53 from desferrioxamine-treated cells with an irrelevant antibody (an IgG2 α control antibody) did not result in co-precipitation of HIF-1 α (Fig. 3b, lane 1). Similarly, immunoprecipitation of p53 from untreated MCF7 cells did not result in HIF-1 α co-precipitation (Fig. 3b, lane 4). Finally, although desferrioxamine induces HIF-1 α accumulation to the same level in SKBr3 and MCF7 cells (Fig. 2), the protein did not co-precipitate with the mutated p53 expressed in SKBr3 cells (data not shown).

We transiently transfected p53-null PC3M cells with expression plasmid pCMV β -wild-type p53, alone or with pCMV β -HA-HIF-1 α (Fig. 3c, d). Although the total DNA concentration was held constant, as well as the absolute amounts of p53- and HIF-1 α -expression plasmids used, co-transfection of both HIF-1 α and p53 reproducibly resulted in a moderate increase in steady-state levels of p53, as measured by western blotting 16–24 h after transfection (in four of five experiments; Figs 3c, 4b). Most important, p53 was easily detected in anti-HA immunoprecipitates obtained from PC3M cells transfected with both HA-HIF-1 α and wild-type-p53 plasmids, but not from cells transfected only with wild-type p53 (Fig. 3d; compare lanes 1–3). Although cells transfected only with pCMV β -wild-type p53 expressed copious amounts of p53 (Fig. 3c, lane 2), anti-HA immunoprecipitation in these cells did not result in co-precipitation of any p53 protein in the absence of co-transfection with HA-HIF-1 α (Fig. 3d, lane 2). Similarly, anti-HA immunoprecipitation in cells transfected with only HA-HIF-1 α did not result in co-precipitation of spurious protein that might have reacted with the anti-p53 antibody used (see Fig. 3d, lane 4).

Finally, we directly assessed the effect of transfected HIF-1 α protein on the p53-responsive reporter plasmid PG13-luciferase in normoxic cells. As expected, HIF-1 α did not directly stimulate PG13 in p53-null PC3M cells, but did augment p53-dependent stimulation by about twofold (Fig. 4a). In MCF7 cells expressing wild-type p53, transient transfection with HIF-1 α augmented basal PG13 activity by about threefold (Fig. 4a). The amount of stimula-

tion seen here is similar to that seen in a previous study, which described a 1.8–2.8-fold induction of reporter activity responsive to wild-type p53 following exposure to hypoxia for 8–18 h (ref. 1). At the same time, co-transfection of HIF-1 α with wild-type p53 into normoxic PC3M cells, or transfection of HIF-1 α alone in normoxic MCF7 cells, moderately increased the steady-state level of p53 as determined by western blotting (Fig. 4b).

Our data indicate, therefore, that induction of HIF-1 α accumulation in response to hypoxic stimuli may be necessary for the concomitant stabilization-dependent accumulation of transcriptionally active wild-type p53, which occurs through direct association of the two proteins. □

Methods

Cells and transfections. MCF7, SKBr3 and Hepa1c1c7 cells were from ATCC. PC3M cells were from J. Trepel, and Hepa1c4 cells were from O. Hankinson. Normal and HIF-1 β ^{-/-} murine ES cells were isolated and cultured as previously described (ref. 11). MCF7 and PC3M cells (7.5×10^5) were transfected with 1 μ g each of plasmid pCMV β -wild-type p53 (from B. Vogelstein) and pCMV β -HA-HIF-1 α (from D. Livingston) in the presence of lipofectamine (GIBCO/BRL), following manufacturer's instructions. Equal amounts of total DNA (2 μ g) were transfected in all experiments. When either of the above plasmids was transfected separately, 1 μ g of pCMV β -galactosidase (Clontech) was included to bring the total DNA transfected to 2 μ g. Cells were lysed and processed for western blotting or immunoprecipitation 16–24 h after transfection. The proteasome inhibitor N-acetyl-Leu-Leu-norleucinal (ALLN), as well as desferrioxamine and cobalt chloride, were from Sigma. These compounds were dissolved in distilled water or dimethylsulphoxide (for ALLN) and used at final concentrations of 100 μ M, 260 μ M and 100 μ M, respectively.

Western blotting and immunoprecipitations. Cells were lysed and nuclear extracts were prepared as described¹⁹. p53 was detected in 30- μ g nuclear extract or total lysate using Ab421 (1:1000, Oncogene Science) as described²⁰. HIF-1 α was detected in 60- μ g nuclear extract using a cocktail of two monoclonal antibodies^{21,22} (OZ12 and OZ15; 1:10 dilution of tissue-culture supernatants, from D. Livingston). p53 or HA-HIF-1 α was immunoprecipitated from 3-mg nuclear-extract protein as previously described²⁰, using 2 μ g of either Ab421 (for p53) or anti-HA antibody (BabCo). p53 was detected in HA immunoprecipitates with anti-p53 antibody Ab7 (Oncogene Science), following manufacturer's suggestions. All blots were developed with SuperSignal chemiluminescence substrate and enhancer solutions (Pierce).

Determination of half-life of p53. The half-life of p53 protein was monitored using pulse-chase analysis with [³⁵S]methionine (L-methionine, 46 Tbq per mmol, ICN), as previously described²⁰. Cells were exposed to either cobalt chloride or desferrioxamine for 6 h before analysis, and the drugs were maintained in the cultures during the half-life determinations.

Exposure of cells to hypoxia. Cells were placed in a hypoxia chamber (Modular Incubator Chamber, Billups-Rothenberg), through which a mixture of 95% N₂ and 5% CO₂ gas was passed at 2 p.s.i. for 2 h. The chamber was then sealed and placed at 37 °C for an extra 6 h, after which time cells were lysed as above.

Northern blotting. Total RNA was prepared from MCF7 cells 6 h after treatment with either cobalt chloride (100 μ M) or desferrioxamine (260 μ M), and from untreated cells. RNA (20 μ g) was northern-blotted with a [³²P]p53 plasmid (pCMV β p53). After analysis, blots were stripped and reprobed with a [³²P]beta-actin plasmid (American Type Culture Collection).

Luciferase assay. 1 μ g of p53-dependent PG13-luciferase reporter plasmid²³ was transfected into PC3M, MCF7, Hepa1c1c7 and Hepa1c4 cells as above, either alone or with pCMV β -wild-type p53 and/or pCMV β -HIF-1 α plasmids (1 μ g each). Total DNA transfected was maintained at 3 μ g by using the pCMV β -galactosidase plasmid as described above. Luciferase activity was monitored in 30 μ g cell extract 16 h after transfection using luciferase assay reagent (Promega).

In experiments testing PG13-luciferase activity in MCF7, Hepa1c1c7, Hepa1c4, normal ES, and ES HIF⁻ cells, 1 μ g of either PG13-luciferase plasmid or pGL2-control plasmid (Promega) was transfected as above and luciferase activity was determined 16 h later. PG13-generated light output was normalized to the light output obtained with pGL2-control in each cell line tested.

Received 3 October; accepted 12 December 1997.

1. Graeber, T. G. *et al.* Hypoxia induces accumulation of p53 protein, but activation of a G1-phase checkpoint by low-oxygen conditions is independent of p53 status. *Mol. Cell. Biol.* **14**, 6264–6277 (1994).
2. Goldberg, M. A., Dunning, S. P. & Bunn, H. F. Regulation of the erythropoietin gene: evidence that the oxygen sensor is a heme protein. *Science* **242**, 1412–1415 (1988).
3. Salceda, S., Beck, I. & Caro, J. Absolute requirement of aryl hydrocarbon receptor nuclear translocator protein for gene activation by hypoxia. *Arch. Biochem. Biophys.* **384**, 389–394 (1996).
4. Gradin, K. *et al.* Functional interference between hypoxia and dioxin signal transduction pathways: competition for recruitment of the Arnt transcription factor. *Mol. Cell Biol.* **16**, 5221–5231 (1996).
5. Guillemain, K. & Krasnow, M. A. The hypoxic response: huffing and HIFing. *Cell* **89**, 9–12 (1997).
6. Wang, G. L. & Semenza, G. L. Desferrioxamine induces erythropoietin gene expression and hypoxia-inducible factor 1 DNA-binding activity: implications for model of hypoxia signal transduction. *Blood* **82**, 3610–3615 (1993).
7. Wang, G. L. & Semenza, G. L. Characterization of hypoxia-inducible factor 1 and regulation of DNA binding activity by hypoxia. *J. Biol. Chem.* **268**, 21513–21518 (1993).
8. Minchenko, A., Salceda, S., Bauer, T. & Caro, J. Hypoxia regulatory elements of the human vascular endothelial growth factor gene. *Cell Mol. Biol. Res.* **40**, 35–39 (1994).
9. Levy, A. P., Levy, N. S., Wegner, S. & Goldberg, M. A. Transcriptional regulation of the rat vascular endothelial growth factor gene by hypoxia. *J. Biol. Chem.* **270**, 13333–13340 (1995).
10. Hoffman, E. C. *et al.* Cloning of a factor required for activity of the Ah (dioxin) receptor. *Science* **252**, 954–958 (1991).
11. Maltepe, E., Schmidt, J. V., Baunoch, D., Bradfield, C. A. & Simon, M. C. Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT. *Nature* **386**, 403–407 (1997).
12. Wang, G. L. & Semenza, G. L. General involvement of hypoxia-inducible factor 1 in transcriptional response to hypoxia. *Proc. Natl Acad. Sci. USA* **90**, 4304–4308 (1993).
13. Wang, G. L., Jiang, B.-H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O₂ tension. *Proc. Natl Acad. Sci. USA* **92**, 5510–5514 (1995).
14. Wang, G. L. & Semenza, G. L. Purification and characterization of hypoxia-inducible factor 1. *J. Biol. Chem.* **270**, 1230–1237 (1995).
15. Wood, S. M., Gleadle, J. M., Pugh, C. W., Hankinson, O. & Ratcliffe, P. J. The role of the aryl hydrocarbon receptor nuclear translocator (ARNT) in hypoxic induction of gene expression. *J. Biol. Chem.* **271**, 15117–15123 (1996).
16. Kallio, P. J., Pongratz, L., Gradin, K., McGuire, J. & Poellinger, L. Activation of hypoxia-inducible factor 1α: posttranscriptional regulation and conformational change by recruitment of the Arnt transcription factor. *Proc. Natl Acad. Sci. USA* **94**, 5667–5672 (1997).
17. Maki, C. G., Huibregtse, J. & Howley, P. M. *In vivo* ubiquitination and proteasome-mediated degradation of p53. *Cancer Res.* **56**, 2649–2654 (1996).
18. Whitesell, L. *et al.* Geldanamycin-stimulated destabilization of mutated p53 is mediated by the proteasome *in vivo*. *Oncogene* **14**, 2809–2816 (1997).
19. Andrews, N. & Faller, D. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res.* **19**, 2499 (1991).
20. Blagosklonny, M., Toretzky, J. & Neckers, L. Geldanamycin selectively destabilizes and conformationally alters mutated p53. *Oncogene* **11**, 933–939 (1995).
21. Arany, Z. *et al.* An essential role for p300/CBP in the cellular response to hypoxia. *Proc. Natl Acad. Sci. USA* **93**, 12969–12973 (1996).
22. Huang, L. E., Arany, Z., Livingston, D. M. & Bunn, H. F. Activation of hypoxia-inducible transcription factor depends primarily upon redox-sensitive stabilization of its α subunit. *J. Biol. Chem.* **271**, 32253–32259 (1996).
23. El-Deiry, W. S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).

Acknowledgements. We thank J. Trepel and O. Hankinson for cell lines, and B. Vogelstein and D. Livingston for plasmids and antibodies.

Correspondence and requests for materials should be addressed to L.N. (e-mail: len@helix.nih.gov).

Enhanced responses to a DNA vaccine encoding a fusion antigen that is directed to sites of immune induction

J. S. Boyle*, J. L. Brady* & A. M. Lew†*

* Autoimmunity and Transplantation Division, Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria 3050, Australia

† Co-operative Research Centre for Vaccine Technology, Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia

Viral infection and vaccination with DNA both induce similar immune responses to encoded antigens that are produced by the host^{1,2}. The availability of antigens in lymphoid organs is important in generating an immune response to viral challenge³. Antigen availability may also be important in the response to DNA vaccines, because immune responses are stronger when antigen is secreted from DNA-transfected cells^{4,5}. We directed antigen to lymphoid organs by vaccination with DNA encoding antigen–ligand fusion proteins. The two ligands examined bind to recep-

tors that are present on high endothelial venule cells of lymph nodes or on antigen-presenting cells. Here we show that both the humoral and the cellular immune responses to a model DNA vaccine were enhanced using either antigen-targeting strategy. Moreover, directing antigen to antigen-presenting cells speeded up, and altered the form of, the immune response. Directing antigen to sites of immune-response induction may represent a generic means of tailoring a potent and effective immune response to a DNA vaccine.

To target antigen to sites of immune induction and thereby increase the effective dose of antigen, human IgG was fused with one of two ligands, L-selectin or cytotoxic T-lymphocyte antigen 4 (CTLA4). L-selectin is expressed on the surface of naive lymphocytes and, by binding to CD34 on high endothelial venule cells, initiates lymphocyte entry into lymph nodes^{6,7}. CTLA4 on activated T cells binds to B7-expressing cells, including antigen-presenting cells (APCs)⁸, which are potent initiators of immune responses. To target our model antigen (human IgG) to the lymph nodes and APCs, we made expression vectors that encoded human IgG fused with the extracellular domains of L-selectin (producing L-SELIg) or CTLA4 (producing CTLA4Ig), respectively.

We first transfected cells with plasmids encoding secreted forms of human IgG (hIg), L-SELIg or CTLA4Ig (Fig. 1a). After purifying supernatants with protein A, we showed by western blotting that all three products were secreted and formed dimers (Fig. 1b); this would enhance binding of these proteins to their respective ligands. Protein expression after DNA immunization was not detected by western blotting of muscle homogenates or sera from B-cell-deficient immunized mice, even after enrichment with protein A (data not shown, and ref. 1).

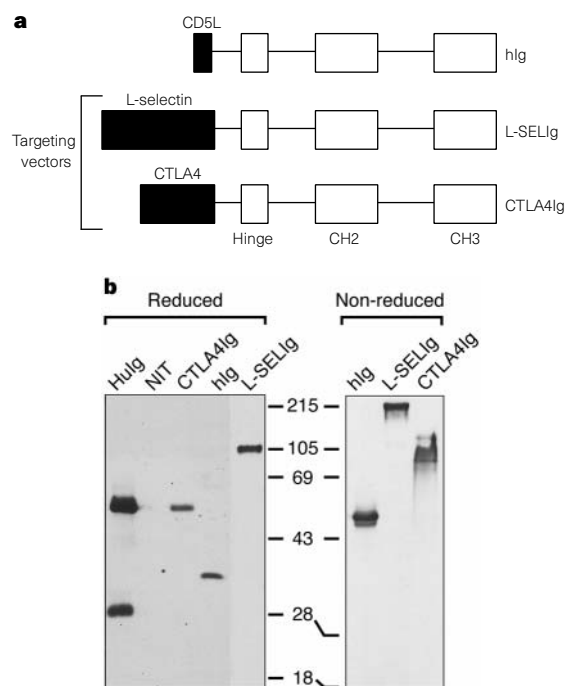


Figure 1 Secretion and dimer formation of human IgG (hlg) and fusion molecules from transfected NIT cells. **a**, Plasmids used for transfections and DNA immunization. The three vectors hlg, L-SELIg and CTLA4Ig encode the hinge, CH2 and CH3 domains of human IgG1 (exons are shown as open boxes). The leader sequence from CD5 (CD5L) that is needed for secretion or the entire extracellular domains of L-selectin or CTLA4 were fused amino-terminal to hlg. **b**, NIT cells were transfected with plasmids encoding the indicated protein. Commercial human immunoglobulin (Hulg) and supernatant from non-transfected NIT cells served as the positive and negative controls, respectively. Non-reduced proteins are dimers.

Specific antibody levels in BALB/c mice were 10,000-fold higher in those mice immunized with CTLA4Ig DNA at 2 weeks (Fig. 2a). At 8 weeks, antibody levels had reached a plateau and were 1,000- and 75-fold higher when CTLA4Ig and L-SELIg, respectively, were used than in controls using hIg (Fig. 2a). Similar results were obtained in repeated experiments with BALB/c and CBA mice. Targeting antigen to lymph nodes or APCs by using L-selectin or CTLA4 also enhanced the T-cell proliferative response. Proliferation was increased by two- to threefold after immunization with L-SELIg, whereas an increase of up to eightfold was achieved with CTLA4Ig (Fig. 2b). Splenocytes from uninjected control mice did not show significant proliferation (data not shown) and background proliferative responses (without antigen) were similar for

all groups. Therefore, our strategy of targeting antigens probably increased both the humoral and the cellular immune responses after immunization with DNA.

A predominance of IgG2a production has been reported after DNA immunization⁹ and this mimics the humoral immune response to viral infection¹⁰. IgG2a and IgG1 differ in function (for example, they differ in their ability to fix complement) and they have been used as indicators of the induction of Th1 and Th2 responses, respectively (for review, see ref. 11). As subclasses of antibodies may be important in vaccine efficacy¹¹, we examined the antibody subclasses in DNA-immunized mice. Levels of proteins in all IgG subclasses (except IgG3, which was not detected) were elevated when the targeting molecules were used (Fig. 2c). We

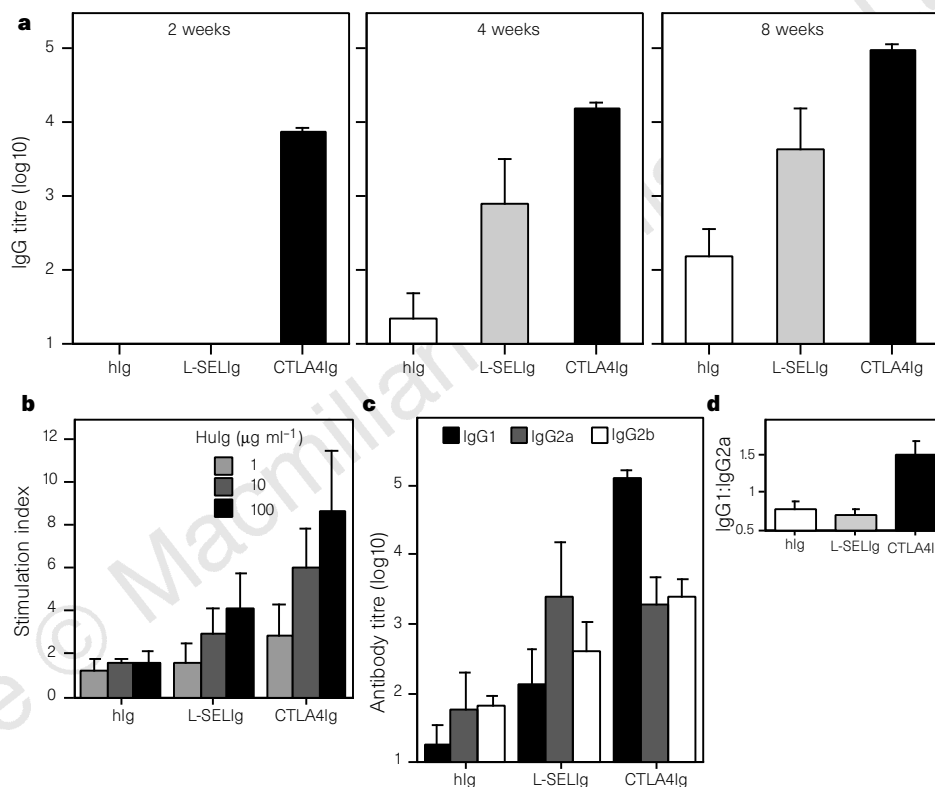


Figure 2 Immune responses to DNA-encoded antigen in BALB/c mice. **a**, The mean Hulg-specific IgG responses $\pm$ SD from three mice in each group is shown after incubation with three different concentrations of Hulg. **b**, The IgG1, IgG2a and IgG2b components of the total IgG responses (**a**) at 8 weeks after

immunization are shown. Hulg-specific IgG3 was not detected. **d**, The mean ratio of the log₁₀ IgG1 to log₁₀ IgG2a titre $\pm$ s.e.m. for individual mice (**c**) at 8 weeks after immunization is shown.

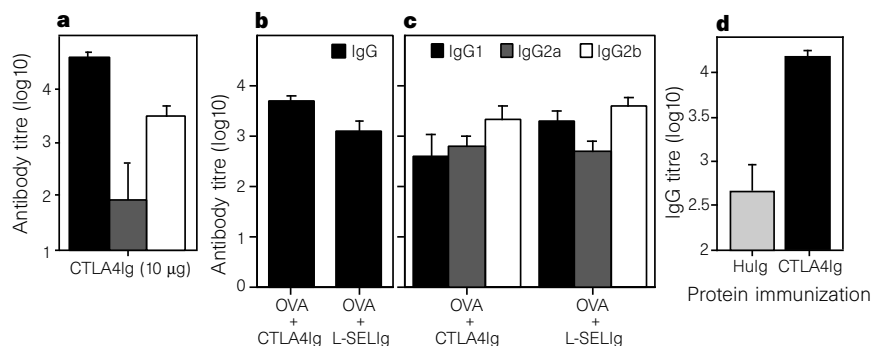


Figure 3 Influence of CTLA4Ig on the antibody response after immunization with DNA or protein. **a**, Levels of the Hulg-specific IgG subclass $\pm$ s.e.m. were determined in 5 mice in each group at 8 weeks after initial immunization with 10 µg of CTLA4Ig vector. Patterns define subclass as in **c**. **b**, **c**, BALB/c mice were injected with 50 µg of each indicated plasmid and the IgG (**b**) or IgG subclass (**c**)

response $\pm$ s.e.m. was determined at 8 weeks after initial immunization. OVA, ovalbumin. **d**, Groups of 10 BALB/c mice were immunized intramuscularly with 5 µg of either Hulg or CTLA4Ig protein purified, using protein A, from the supernatant of transfected NIT cells. Sera were obtained 2 weeks after immunization and the Hulg-specific IgG response $\pm$ s.e.m. was determined.

compared the IgG1:IgG2a ratio in individual mice; the predominance of IgG2a and IgG2b in L-SELg-immunized mice was similar (although greater) to hIg controls, whereas the response to CTLA4Ig was different (Fig. 2d) as a 7,000-fold increase in IgG1 levels was seen (Fig. 2c).

To determine whether the antibody response elicited with CTLA4Ig was simply greater and therefore dominated by IgG1, we immunized mice with suboptimal doses of CTLA4Ig so that the subclass dominance in mice with lower total IgG levels could be compared. BALB/c mice were immunized by intramuscular injection of suboptimal doses (1, 5, 10, 25 or 50 µg) of DNA encoding CTLA4Ig on days 0 and 14. Levels of the whole human immunoglobulin (HuIg)-specific IgG subclasses were determined 8 weeks after initial immunization. The predominance of IgG1 was seen in all mice receiving 50 µg, 25 µg or 10 µg of CTLA4Ig (Fig. 3a; results of use of 10 µg are shown) and in the responding mice that received doses of <10 µg (not all mice had detectable IgG responses). The IgG1 predominance suggested that CTLA4 might have caused a non-specific change in the immune response, possibly by direct stimulation of APCs. Indeed, modulation of the immune response has been shown to occur as a result of CTLA4-mediated blockage of costimulation by interaction of CD28 and B7, but in this case there is an immunosuppressive effect and a reduction of the antibody response to co-administered sheep red blood cells and keyhole-limpet haemocyanin¹². However, high doses of CTLA4 (50 µg) are required to reduce levels of available surface B7 enough to block stimulation with CD28 and achieve *in vivo* immune suppression^{12,13}; such doses are well above (10,000-fold) that achieved with DNA injection¹⁴. At lower doses, CTLA4 enhances the immune response to fused antigens by delivering the antigens to APCs and, therefore, to lymphoid organs. In our system, an immunosuppressive effect of CTLA4 was unlikely as the CTLA4Ig fusion enhanced both the cellular and the humoral immune responses. We obtained direct evidence that CTLA4 did not cause any nonspecific modulation of the immune response when we immunized mice with CTLA4Ig and DNA encoding ovalbumin. The ovalbumin-specific IgG and IgG-subclass titres in these mice were similar to those in control mice (Fig. 3c). The HuIg-specific response of IgG and IgG subclasses was also determined and was similar to that shown previously (Fig. 2a, c). This indicates that the increases in IgG and IgG1 levels achieved with CTLA4 were not due to nonspecific immunomodulation and required direct fusion of antigen with the targeting ligand.

The increased antibody responses were not only achieved after immunization with DNA. We immunized mice by intramuscular injection of soluble whole HuIg or CTLA4Ig proteins purified from a transfected cell line (Fig. 1b). Mice that received a single injection of the soluble CTLA4Ig had significantly higher antibody responses compared with mice injected with HuIg (Fig. 3d). Increased immune responses, with elevated IgG1 levels, were not only seen in response to HuIg. Targeting vectors containing CTLA4 DNA were found to induce high antibody levels, with similar rapid kinetics, against anti-Igh-1^b allotype sera for use as a laboratory reagent and against 45W, a vaccine antigen against cysticercosis¹⁵. These results suggest that antigen targeting with CTLA4 is a generic approach that enhances antibody responses in the absence of adjuvant.

Intramuscular¹⁴ or intradermal² injection of antigen encoding plasmid DNA induces both cellular and humoral immune responses and holds great vaccine potential¹. Potency of immune responses will be important in vaccine efficacy. Previously, an increase in the antibody response to an encoded antigen was achieved by co-injection of a plasmid that expressed granulocyte-macrophage colony-stimulating factor¹⁶, presumably because of increased recruitment of APCs to the site of antigen expression. We, however, targeted DNA-encoded antigen to APCs and lymph nodes directly to increase cellular and humoral immune responses. Our results indicate that, by targeting different lymphoid compartments, some

control over the magnitude and form of the immune response may be gained. The model system we describe may be useful in the development of both genetic and more conventional vaccines so that the immune response most likely to confer protection can be generated. □

Methods

Plasmids and immunizations. The hIg vector containing the CD5 leader sequence for secretion and the L-SELg vector were derived from clones provided by B. Seed and were inserted into the Rep10 plasmid (Invitrogen). The complementary DNA encoding the CTLA4 fusion protein (CTLA4Ig; provided by P. Lane) was inserted into the Rep7 plasmid (Invitrogen). Rep7 and Rep10 differ only in the orientation of the multiple cloning site and contain a Rous sarcoma virus promoter, a simian virus 40 intron and a polyadenylation signal. To examine protein expression, NIT (mouse insulinoma) cells were transfected with plasmids encoding human IgG, L-SELg or CTLA4Ig. Secreted proteins were detected by western blotting after purification of supernatants using protein A¹⁷. Plasmids for injection were prepared by precipitation with polyethylene glycol^{14,18}. Endotoxin was removed from plasmid preparations by four Triton X-114 phase separations¹⁹ and the resulting plasmid preparations contained less than 0.05 ng mg⁻¹ DNA, as determined by the limulus amoebocyte lysate assay (BioWhittaker). Female mice aged 6 to 8 weeks received 100 µg of plasmid DNA or the indicated quantity of soluble protein in normal saline intramuscularly in both quadriceps, on days 0 and 14 of each experiment unless otherwise stated.

Immunological assays. Human immunoglobulin- or ovalbumin-specific IgG and IgG subclass responses were determined by enzyme-linked immunosorbent assays¹⁸. Briefly, microtitre plates (Dynatech) coated with whole HuIg protein (Intragam, Melbourne; 10 µg ml⁻¹ in PBS) or ovalbumin (Sigma; 10 µg ml⁻¹ in PBS) were incubated with serially diluted sera in blocking buffer (5% skimmed-milk powder in PBS) overnight at 4 °C. Bound antibody was detected after incubation with peroxidase-conjugated antibodies to mouse IgG, IgG1, IgG2a, IgG2b or IgG3 (Southern Biotechnology) diluted in blocking buffer. Titres were defined as the highest dilution to reach an A₄₅₀ of 0.2. The proliferation of 2 × 10⁵ splenocytes was determined by a standard 5-day ³H-thymidine-uptake protocol at 6 weeks after initial immunization. The mean stimulation index was calculated as the c.p.m. of splenocytes with antigen divided by c.p.m. of splenocytes alone.

Received 18 September; accepted 11 December 1997.

- Ulmer, J. B. *et al.* Heterologous protection against influenza by injection of DNA encoding a viral protein. *Science* **259**, 1745–1749 (1993).
- Raz, E. *et al.* Intradermal gene immunization: the possible role of DNA uptake in the induction of cellular immunity to viruses. *Proc. Natl Acad. Sci. USA* **91**, 9519–9523 (1994).
- Karrer, U. *et al.* On the key role of secondary lymphoid organs in antiviral immune responses studied in lymphoplastic (Aly/Aly) and spleenless (Hox11(–/–)) mutant mice. *J. Exp. Med.* **185**, 2157–2170 (1997).
- Boyle, J. S., Konias, C. & Lew, A. M. Influence of cellular location of expressed antigen on the efficacy of DNA vaccination: cytotoxic T lymphocytes and antibody responses are sub-optimal when antigen is cytoplasmic after intramuscular DNA immunization. *Int. Immunol.* **9**, 1897–1906 (1997).
- Inchausti, G. *et al.* Plasmid DNA expressing a secreted or a nonsecreted form of hepatitis C virus nucleocapsid: comparative studies of antibody and T-helper responses following genetic immunization. *DNA Cell Biol.* **16**, 185–195 (1997).
- Arbones, M. L. *et al.* Lymphocyte homing and leukocyte rolling and migration are impaired in L-selectin-deficient mice. *Immunity* **1**, 247–260 (1994).
- Catalina, M. D. *et al.* The route of antigen entry determines the requirement for L-selectin during immune responses. *J. Exp. Med.* **184**, 2341–2346 (1996).
- Linsley, P. S., Clarke, E. A. & Ledbetter, J. A. T-cell antigen CD28 mediates adhesion with B cells by interacting with activation antigen B7/BB-1. *Proc. Natl Acad. Sci. USA* **87**, 5031–5035 (1990).
- Mor, G. *et al.* Complexity of the cytokine and antibody response elicited by immunizing mice with *Plasmodium yoelii* circumsporozoite protein plasmid DNA. *J. Immunol.* **155**, 2039–2046 (1995).
- Michel, M. L. *et al.* DNA-mediated immunization to the hepatitis B surface antigen in mice: aspects of the humoral response mimic hepatitis B viral infection in humans. *Proc. Natl Acad. Sci. USA* **92**, 5307–5311 (1995).
- Snapper, C. M. & Mond, J. J. Towards a comprehensive view of immunoglobulin class switching. *Immunol. Today* **15**, 15–17 (1993).
- Linsley, P. *et al.* Immunosuppression *in vivo* by a soluble form of the CTLA-4 T cell activation molecule. *Science* **257**, 792–795 (1992).
- Lenschow, D. J. *et al.* Long-term survival of xenogeneic pancreatic islet grafts induced by CTLA4Ig. *Science* **257**, 789–792 (1992).
- Wolff, J. A. *et al.* Direct gene transfer into mouse muscle *in vivo*. *Science* **247**, 1465–1468 (1990).
- Johnson, K. S. *et al.* Vaccination against ovine cysticercosis using a defined recombinant antigen. *Nature* **338**, 585–587 (1989).
- Xiang, Z. & Ertl, H. C. Manipulation of the immune response to a plasmid-encoded viral antigen by coinoculation with plasmids expressing cytokines. *Immunology* **2**, 129–135 (1995).
- Lew, A. M., Brady, J. L., Silva, A., Coligan, J. E. & Georgiou, H. M. Secretion of CTLA4Ig by an SV40 T antigen transformed islet cell line inhibits graft rejection against the neoantigen. *Transplantation* **62**, 83–89 (1996).

18. Boyle, J. S., Silva, A., Brady, J. L. & Lew, A. M. DNA immunization: induction of high avidity antibody and effect of route on T cell cytotoxicity. *Proc. Natl Acad. Sci. USA* **94**, 14626–14631 (1997).
19. Aida, Y. & Pabst, M. J. Removal of endotoxin from protein solutions by phase separation using Triton X-114. *J. Immunol. Meth.* **132**, 191–195 (1990).

Acknowledgements. We thank J. F. A. P. Miller and L. C. Harrison for help in preparing this manuscript, and W. R. Heath, R. M. Sutherland, D. M. Tarlinton, D. Drew and R. Strugnell for helpful discussions. This work was supported by NH and MRC of Australia, Appel Estate, Diabetes Australia and a JDFI-NIH grant.

Correspondence and requests for materials should be addressed to A.M.L. (e-mail: lew@wehi.edu.au).

Rhodopsin-family receptors associate with small G proteins to activate phospholipase D

Rory Mitchell, Derek McCulloch, Eve Lutz, Melanie Johnson, Chris MacKenzie, Myles Fennell, George Fink, Wei Zhou* & Stuart C. Sealfon*

MRC Brain Metabolism Unit, 1 George Square, Edinburgh, EH8 9JZ, UK

*Fishberg Center for Neurobiology and Department of Neurology, Mount Sinai School of Medicine, New York, New York 10029, USA

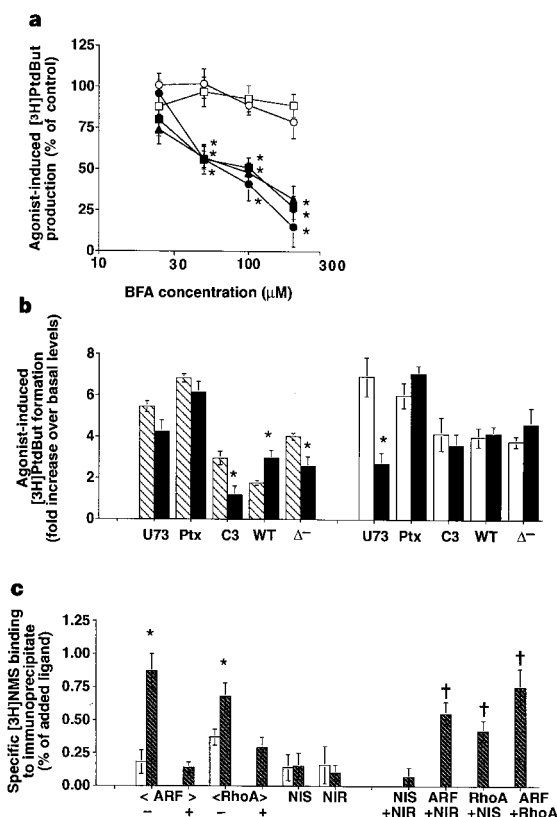
G-protein-coupled receptors of the rhodopsin family transduce many important neural and endocrine signals. These receptors activate heterotrimeric G proteins and in many cases also cause activation of phospholipase D, an enzyme that can be controlled by the small G proteins ARF and RhoA^{1–3}. Here we show that the activation of phospholipase D that is induced by many, but not all, Ca²⁺-mobilizing G-protein-coupled receptors is sensitive to inhibitors of ARF and of RhoA. Receptors of this type were co-immunoprecipitated with ARF or RhoA on exposure to agonists, and the effects of GTP analogues on ligand binding to the receptor changed to a profile that is characteristic of small G proteins. These receptors contain the amino-acid sequence Asn-ProXXTyr in their seventh transmembrane domain, whereas

receptors capable of activating phospholipase D without involving ARF contain the sequence AspProXXTyr. Mutation of this latter sequence to AsnProXXTyr in the gonadotropin-releasing hormone receptor conferred sensitivity to an inhibitor of ARF, and the reciprocal mutation in the 5-HT_{2A} receptor for 5-hydroxytryptamine reduced its sensitivity to the inhibitor. Receptors carrying the AsnProXXTyr motif thus seem to form functional complexes with ARF and RhoA.

The activation of phospholipase D (PLD) (measured as formation of [³H]phosphatidylbutanol) by a number of receptors that are native to 1321N1 human astrocytoma cells⁴ was differentially inhibited by brefeldin A (BFA), an inhibitor of guanine-nucleotide exchange on ARF⁵. Activation of PLD by H₁ histamine, B₂ bradykinin and M₃ muscarinic receptors in these cells⁶ (as for M₃ receptors expressed in HEK 293 cells⁷) was sensitive to BFA whereas activation of PLD by thrombin and thromboxane A₂ receptors was BFA-resistant (Fig. 1a). Thrombin- but not M₃-receptor responses were attenuated by the phospholipase C (PLC) inhibitor U73122, but neither response was affected by pertussis toxin (Fig. 1b). The Rho inhibitor C3 exoenzyme and a negative construct, CMV5 Asn 19 RhoA (ref. 8), both reduced M₃- but not thrombin-receptor-mediated activation of PLD, whereas wild-type RhoA increased responses to a lower concentration of carbachol but not thrombin (Fig. 1b). These results indicate that some Ca²⁺-mobilizing G-protein-coupled receptors (GPCRs) use a pathway of PLD activation that is independent of Gq/11 (which activates PLC) and Gi/o, yet involves ARF and RhoA.

We used co-immunoprecipitation to test whether M₃-receptor-mediated PLD activation might involve a step in which the receptor and small G proteins interact closely. Figure 1c shows that solubilized M₃ receptors (levels of which were measured by binding of [³H]N-methyl scopolamine ([³H]NMS)) could be immunoprecipitated using polyclonal antibodies against ARF1/3 (ref. 9) or RhoA¹⁰. Yields from using combinations of antibodies were less than additive. Co-immunoprecipitation required pre-exposure to agonists (priming) and there was little co-immunoprecipitation when

Figure 1 Properties of agonist-evoked PLD responses in 1321N1 cells and co-immunoprecipitation of M₃ receptors with ARF1/3 and RhoA antibodies. **a**, the effects of BFA on [³H]phosphatidylbutanol ([³H]PtdBut) production elicited by: ● 200 μ M carbachol, ■ 10 μ M bradykinin, ▲ 2 mM histamine, ○ 0.5 units per ml thrombin, or □ 30 μ M U46619 (a TXA₂-receptor agonist). IC₅₀ values for effects of BFA on M₃, bradykinin and histamine responses were 72 $\pm$ 11, 78 $\pm$ 15 and 82 $\pm$ 13 μ M, respectively, whereas thrombin-receptor and TXA₂-receptor responses were resistant up to 200 μ M BFA. **b**, Production of PtdBut in response to carbachol and thrombin is shown by hatched and open columns, respectively. Corresponding values for reagent-treated cells are shown by adjacent black columns. Reagents were U73122 (20 μ M; U 73), pertussis toxin (200 ng ml⁻¹ for 18 h; Ptx), C3 exoenzyme (4.8 μ g ml⁻¹; C3) and wild-type or negative RhoA constructs (WT and Δ^-). Carbachol was used at a concentration of 200 μ M except in C3 and WT RhoA experiments, where 100 μ M and 20 μ M, respectively, were used. Thrombin was usually 0.5 units per ml but 0.2 units per ml in WT RhoA experiments. In **a**, **b**, values are means $\pm$ s.e.m.; n = 4–10; significant changes from the control levels are indicated by asterisks (P < 0.05; Wilcoxon test). **c**, Solubilized membrane proteins from cells preincubated with carbachol (hatched columns) or from controls (open columns) were immunoprecipitated with sheep anti-ARF1/3 antibody (ARF), rabbit anti-RhoA antibody (RhoA), non-immune sheep IgG (NIS) or non-immune rabbit IgG (NIR) or combinations of these antibodies, before labelling M₃ receptors with [³H]NMS. In some ARF/RhoA immunoprecipitations, blocking peptides were included (shown as $\pm$). Values are means $\pm$ s.e.m. (n = 6–9). For asterisked bars, P < 0.05 compared with unprimed controls, and for bars with daggers P < 0.05 compared with non-immune IgG controls (Wilcoxon test). We estimate that up to 48% and 27% of ligand binding to solubilized M₃ receptor may be associated specifically (in a peptide-blockable manner) with ARF and RhoA immunoprecipitates after agonist priming.



using control immunoreagents or in the presence of excess peptide antigen. Use of BFA (100 μ M) during priming reduced the yield of M_3 receptors in ARF1/3 immunoprecipitates by $74 \pm 19\%$. Thrombin receptors, labelled with [125 I]Ala-pFPhe-Arg-Cha-HArg-Tyr-NH $_2$ ([125 I]TRP; ref. 11) could not be co-immunoprecipitated by antibodies against ARF/Rho after priming with thrombin agonist (data not shown). When a control immunoprecipitating antibody (anti-PKC α , Transduction Laboratories) was used to collect another signalling protein known to translocate to 1321N1 cell membranes, preincubation with carbachol caused no increase in [3 H]NMS binding to immunoprecipitates despite a 3.1 ± 0.3 -fold increase in membrane binding of [3 H]phorbol 12,13-dibutyrate (data not shown).

Immunoblotting using distinct (monoclonal) antibodies showed that authentic ARF and Rho were present in ARF/Rho-directed immunoprecipitates from carbachol-primed cells; ARF and Rho were not present in immunoprecipitates when peptide antigens to the polyclonal reagents were added (Fig. 2a, b). In the detergent/NaCl conditions that we used, the monoclonal anti-ARF antibodies ID9 (ref. 12) and clone 26 (Transduction Laboratories) were much less effective or completely ineffective, respectively, at precipitating ARF immunoreactivity and [3 H]NMS-binding sites (data not shown). An M_3 -receptor antiserum fraction⁶, but not non-immune IgG, caused co-precipitation of immunoreactive ARF and RhoA after priming with carbachol (Fig. 2c, d). Similarly, polyclonal antibodies to the AT $_1$ receptor (BFA inhibited PLD activation by this receptor; half-maximal inhibitory concentration (IC_{50}) $58 \pm 5 \mu$ M) caused priming-dependent, peptide-blockable co-precipitation of authentic ARF (Fig. 2e) and Rho (data not

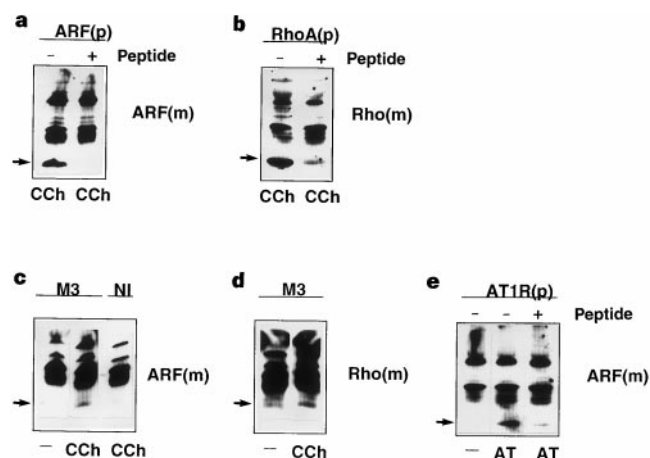


Figure 2 Immunoblots for ARF and Rho on immunoprecipitates generated with polyclonal ARF1/3, RhoA and receptor antibodies. Blots **a-d** are from 1321N1 cells and **e** is from rat anterior pituitary cells. For each panel, the labels below indicate the preincubation conditions (CCh, carbachol; AT, AT $_1$; -, control). Labels above show the immunoprecipitation reagents (ARF(p), sheep polyclonal anti-ARF1/3 antibody; RhoA(p), rabbit polyclonal anti-RhoA antibody; M $_3$, rabbit polyclonal anti-M $_3$ -receptor antibody; NI, non-immune rabbit IgG; AT1R(p), rabbit polyclonal anti-AT $_1$ -receptor antibody), with (+) or without (-) blocking peptides. Labels to the right indicate the immunoblotting reagents (ARF(m), mouse monoclonal anti-ARF antibody; Rho(m), mouse monoclonal anti-Rho antibody). The arrow indicates the position of the 20 kDa molecular mass standard.

shown). In addition, when AT $_1$ -primed membranes were treated with the crosslinker ethylene glycolbis(succinimidylsuccinate) and immunoprecipitations carried out with antibody ID9 in the presence of Triton X-100/SDS, specific AT $_1$ -receptor immunoreactivity became detectable in immunoprecipitated high-molecular-mass complexes of approximately 170K (data not shown). These results indicate that certain receptors may be able to form physical complexes involving ARF1/3 and/or RhoA and that this may be important in the activation of PLD.

We used GTP analogues to determine whether the effects of these analogues on both receptor and effector properties were consistent with the above hypothesis. The affinity of M_3 receptors labelled with [3 H]NMS for carbachol was reduced by guanosine 5'-O-(3-thio)-triphosphate (GTP γ S), BeF $_3$ (an isostere of AlF $_4$ (ref. 13)) and guanosine 5'-[β -methylene]triphosphate (GPPCH $_2$ P) (Table 1a).

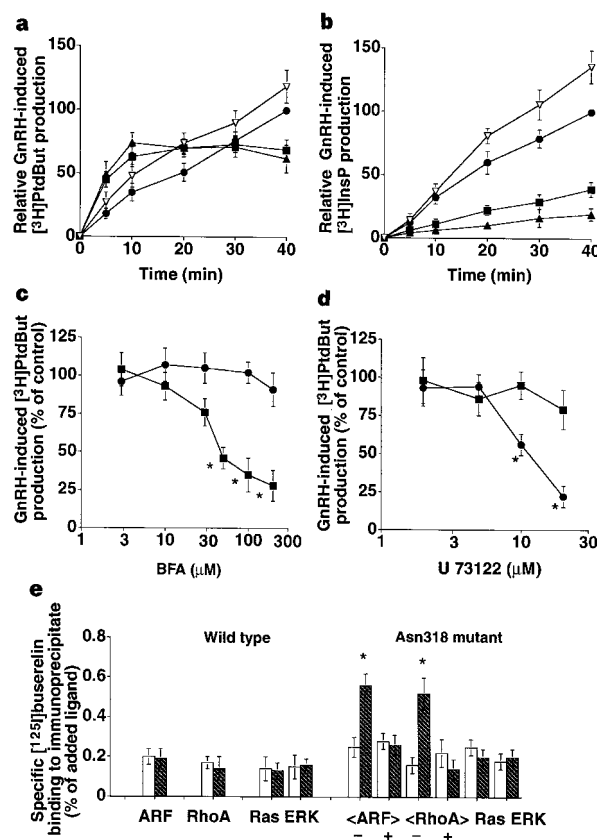


Figure 3 PLD and PLC responses of wild-type, Asn318 and Asp318 plus Asp87 Asn318 mutant GnRH receptors expressed in COS 7 cells and receptor immunoprecipitation with ARF1/3 and RhoA antibodies. **a, b** Show the time course of [3 H]phosphatidylbutanol ([3 H]PtdBut) and [3 H]inositol phosphate ([3 H]InsP) production evoked by 100 nM GnRH at $\bullet$ wild-type, $\blacksquare$ Asn318 mutant and $\blacktriangle$ Asp87 Asn318 mutant receptors and by ∇ AlF $_4$ (10 mM NaF and 30 μ M AlCl $_3$) in Asn318 mutant expressing cells. Values are percentages of wild-type response at 40 min (means $\pm$ s.e.m.; $n = 4-6$). **c, d** Show effects of BFA and U73122 on [3 H]PtdBut production evoked by 100 nM GnRH at $\bullet$ wild-type and $\blacksquare$ Asn318 mutant receptors. Values are means $\pm$ s.e.m.; $n = 4-6$. **e**, Shows binding of [125 I]busserelin to immunoprecipitates from cells expressing wild-type or Asn318 mutant receptors. Extracts from cells preincubated with GnRH (hatched columns) or from controls (open columns) were immunoprecipitated with antibodies against ARF1/3 (ARF), RhoA (RhoA), p21Ras (Ras) or ERK1/2 (ERK). Blocking peptides for ARF/RhoA antibodies were sometimes included (indicated as $\pm$). Values are means $\pm$ s.e.m.; $n = 4-8$. Asterisks relate to P values of <0.05 compared to controls, by Wilcoxon test. After agonist priming, 26% and 33% of ligand binding to solubilized mutant GnRH receptor (1100-1400 c.p.m. per sample) seemed to be associated specifically (in a peptide-blockable manner) with ARF and RhoA immunoprecipitates whereas specific association with these precipitates of the wild-type was negligible.

After carbachol priming, the effect of GTP γ S was unaltered, that of BeF $_3$ was attenuated and that of GPPCH $_2$ P was increased in a way that was $57 \pm 10\%$ inhibited by BFA. Table 1b shows that GTP analogues activated PLD in permeabilized cells and that their effects were modified by agonist priming^{14–16}. After carbachol priming of 1321N1 cells, the effect of GTP γ S did not change significantly but that of BeF $_3$ was reduced and that of GPPCH $_2$ P was greatly increased. Agonist priming lowered the effective concentration for half maximum response (EC $_{50}$) for GPPCH $_2$ P from $367 \pm 74 \mu\text{M}$ to $97 \pm 11 \mu\text{M}$ (6% and 22%, respectively, of the potency of GTP γ S (itself with an unaltered EC $_{50}$ of 21–22 μM)). Agonist exposure causes ARF1/3 and RhoA^{7,16,17} to translocate to cell membranes and the increased effect of GPPCH $_2$ P here was abrogated by BFA or C3 exoenzyme (Table 1b). Effects of AlF $_4^-$ and BeF $_3^-$ may involve trimeric, but not small, G proteins^{13,18} (although selectivity may be modified *in vivo*). In contrast, $\beta\gamma$ -methylene analogues of GTP may retain a substantial fraction of the potency of GTP γ S at ARF, Rab5 and other small G proteins, despite being very weak activators of trimeric G proteins and Ras^{19,20}. Indeed in an assay of [^{35}S]GTP γ S association to native ARF, collected from ID9 immunoprecipitates (which were later renatured) of urea-treated extracts from carbachol-primed 1321N1-cell membranes, GPPCH $_2$ P retained 30% of the potency of GTP γ S with mean IC $_{50}$ values of 54 and 16 μM , respectively (R.M. and D.M., unpublished observations). These results concur with the idea that agonist-induced translocation of ARF/RhoA to membranes may promote their involvement in some form of complex with GPCRs, and may ultimately enhance receptor-mediated activation of PLD.

Each of the receptors that were sensitive to BFA, C3 exoenzyme or CMV5 Asn 19 RhoA or that could be co-immunoprecipitated with ARF 1/3 or RhoA contains the canonical AsnProXXTyr motif (where X represents any amino acid) in transmembrane domain VII (TMD VII), whereas the other receptors contain AspProXXTyr instead. The wild-type gonadotropin-releasing hormone (GnRH) receptor (which contains AspProXXTyr) and two mutants containing an AsnProXXTyr motif (containing mutation of Asp 318 $\rightarrow$ Asn or Asn 87 $\rightarrow$ Asp as well as Asp 318 $\rightarrow$ Asn (ref. 21)) were expressed in COS 7 cells. The wild-type receptor activated PLD linearly over 30–40 min, like the receptor in α T3-1 cells²². In contrast, the Asn 318 and Asp 87 Asn 318 mutants displayed initial rates of

[^3H]phosphatidylbutanol formation that were more than 2.5-fold that of the wild-type receptor (Fig. 3a). Enhanced PLD coupling in the mutants was observed despite reduced [^3H]inositol phosphate production (Fig. 3b) and reduced membrane [^{125}I]buserelin binding (for which values of 589 ± 39 , 230 ± 41 and 148 ± 36 fmol per mg protein B_{max} were obtained for cells expressing wild-type, single mutant and double mutant receptors, respectively). PLD-activation responses of the mutants desensitized rapidly, whereas PLC-activation responses did not. In contrast, postreceptor stimulation of trimeric G proteins with AlF $_4^-$ caused non-desensitizing activation of both PLD and PLC in Asn 318 mutant cells. Neither altered receptor kinetics nor enhanced receptor internalization to allow contact with intracellular PLD²³ could explain the enhanced PLD coupling, as [^{125}I]buserelin association with the membrane and receptor internalization rates in wild-type and Asn 318-mutant cells were indistinguishable. In addition, 200 μM monodansylcadaverine and 30 μM monensin (which caused 70–80% inhibition of receptor internalization) increased rather than reduced, both PLD responses (data not shown).

PLD activation by the wild-type receptor was inhibited by U73122 (IC $_{50}$ of $11 \pm 1 \mu\text{M}$) but not by BFA in concentrations up to 200 μM (these data are similar to data from α T3-1 cells), whereas the Asn 318-mutant response was sensitive to BFA (IC $_{50}$ of $54 \pm 8 \mu\text{M}$) but not U73122 in concentrations up to 20 μM (Fig. 3c, d). ATP, an agonist for the native P $_{2u}$ receptor (which contains the AspProXXTyr motif), and ionomycin/phorbol ester both caused BFA-resistant PLD activation (data not shown). The gain of BFA-sensitive PLD (but not PLC) activation upon replacement of the AspProXXTyr motif with AsnProXXTyr indicates that this structural motif may be important in gating ARF/Rho-mediated coupling to PLD. This hypothesis is supported by the greater sensitivity to BFA (IC $_{50}$ $47 \pm 11 \mu\text{M}$) of PLD activation by the wild-type 5-HT $_{2A}$ receptor (which contains AsnProXXTyr) than its Asp 376 mutant (ref. 24) (24 ± 13 inhibition at 200 μM BFA) expressed in COS 7 cells.

The importance of the AsnProXXTyr motif in the proposed linkage between receptor and small G proteins was directly indicated by co-immunoprecipitation of agonist-treated Asn 318 mutant but not wild-type GnRH receptors using ARF and RhoA antibodies (Fig. 3e). Furthermore, although GTP γ S increased dissociation of [^{125}I]buserelin from both the Asn 318 mutant and the wild-type receptor in permeabilized cells, GPPCH $_2$ P was effective towards the mutant only and the inverse was true for BeF $_3^-$ (Table 1c). The effect of GTP γ S at the mutant, but not the wild-type, receptor was inhibited (by $62 \pm 12\%$) by 50 μM BFA.

Our results suggest a model involving a previously unrecognized association between certain rhodopsin-family receptors and the small G proteins ARF and RhoA. One functional consequence of this appears to be an enhanced coupling of receptors to PLD activation. The interaction seems to occur when an AsnProXXTyr, but not AspProXXTyr, receptor motif is present and may be enhanced by agonist-induced translocation of ARF/RhoA to the plasma membrane. The form and site of interaction of small G proteins with receptors is unknown, but it seems likely that other proteins that act as adapters or regulators of small G protein function^{25–28} may participate or mediate in the association. □

Methods

Phospholipase assays. PLD and PLC activities were monitored as production of [^3H]phosphatidylbutanol and [^3H]inositol phosphates, respectively²⁹. For assays of PLD activity in acutely permeabilized cells, prelabelled 1321N1 cells were primed with 100 μM carbachol (10 min, 37 °C) or control before intracellular buffer⁷ was added. This buffer contained 2 mM MgATP, 10 μM NAD, 6 μM digitonin, 30 mM butan-1-ol, and GTP γ S, GPPCH $_2$ P or F $^-$ (NaF in the presence of 30 μM BeCl $_2$) and BFA or C3 exoenzyme as required. Male rat anterior hemipituitaries were labelled for 2 h *in vitro* for assay of AT $_{11}$ -evoked PLD activation.

Table 1 Effects of GTP analogues on agonist recognition by M $_3$ and GnRH receptors and on PLD activation

Assay and treatment	GTP analogue		
	GTP γ S	F $^-$	GPPCH $_2$ P
(a) Carbachol affinity for [^3H]NMS-binding sites in 1321N1 cells (fold increase in IC $_{50}$)			
Control	5.23 ± 0.97	4.06 ± 0.80	2.31 ± 0.34
Carbachol-primed	4.48 ± 0.34	$1.82 \pm 0.31^*$	$3.46 \pm 0.40^\dagger$
(b) Activation of PLD in permeabilized 1321N1 cells (Increase over basal activation (%))			
Control	150 ± 9	115 ± 14	48 ± 8
+BFA		117 ± 9	$31 \pm 3^\ddagger$
+C3 exoenzyme		95 ± 11	$27 \pm 5^\ddagger$
Carbachol-primed	168 ± 10	$78 \pm 10^*$	$110 \pm 12^\dagger$
+BFA		69 ± 11	$38 \pm 9^\ddagger$
+C3 exoenzyme		85 ± 9	$23 \pm 7^\ddagger$
(c) [^{125}I]buserelin dissociation rate in transfected COS7 cells (Reduction in $t_{1/2}$ of slow component (%))			
Wild-type GnRH receptor	43 ± 4	36 ± 8	12 ± 5
Asn 318 mutant	40 ± 5	$7 \pm 5^*$	$38 \pm 8^\ddagger$

GTP γ S, F $^-$ and GPPCH $_2$ P were present at 100 μM , 10 mM and 200 μM , respectively, except in **c** where 3 mM F $^-$ and 100 μM GPPCH $_2$ P were used. BFA and C3 exoenzyme were used at 100 μM and 4.8 $\mu\text{g ml}^{-1}$, respectively. Values are means $\pm$ s.e.m., $n = 4$ –10. Statistically significant differences ($P < 0.05$ by Wilcoxon test): *, less than corresponding unprimed control; (a, b) or wild-type control (c); † greater than corresponding unprimed or wild-type control; ‡ reversal of GTP-analogue effect.

Liposome treatment and transfection of 1321N1 cells. Cells were treated with lipofectamine containing C3 exoenzyme (2 µg per well, 5 h), CMV5-RhoA constructs (0.5 µg DNA per 4.5 cm² well, 7 h) or control⁸. Rho-construct cells were assayed after 48 h.

Solubilization and immunoprecipitation of [³H]NMS-labelled M₃ receptors and [¹²⁵I]TRP-labelled thrombin receptors. 1321N1 cells (preincubated with carbachol (100 µM) or the thrombin agonist Ser-Phe-Leu-Arg-Asn-NH₂ (30 µM) for 10 min) were homogenized in cold Hepes buffer with peptidase and phosphatase inhibitors. Membranes were solubilized in 5 mM CHAPS, 0.1% Na cholate and 1M NaCl for 30 min at 4 °C. An equal volume of 20% glycerol in CHAPS/cholate buffer without NaCl was added (with 0.6 mg ml⁻¹ phosphatidylcholine for M₃ receptors). Supernatant was precleared and incubated (4 °C, 18 h) with sheep anti-ARF1/3 immunoglobulins (10–15 µl/ml; antigen ARF1_{98–112}; gift from M. J. O. Wakelam)⁹ or with an immunoprecipitating rabbit anti-RhoA IgG (2–3 µg ml⁻¹; antigen RhoA_{119–132}; Santa Cruz Biotechnology)¹⁰. The ARF antiserum immunoprecipitated authentic immunoreactive ARF from CHAPS/cholate/NaCl membrane extracts (Fig. 2a) and from cytosol after its incubation with GTPγS and the addition of CHAPS/cholate/NaCl but not without this treatment (R.M. and M.J., unpublished observations). Blocking peptides were used at 6 µg ml⁻¹ and control non-immune IgG at 3 µg ml⁻¹. Immune complexes were collected with protein-G-Sepharose. M₃ receptors were assayed in 10 mM MgCl₂, 200 mM NaCl, 10% glycerol, 3 mg ml⁻¹ phosphatidylcholine, 20 mM Hepes pH 7.5, 10 nM [³H]NMS (85 Ci mmol⁻¹, Du Pont) with or without 10 µM N-methyl atropine for 40 min at 37 °C, before precipitation with polyethylene glycol. Thrombin receptors were assayed in 10 mM MgCl₂, 150 mM NaCl, 0.25% BSA, 0.05% bacitracin, 0.1 mM 4-(2-aminoethyl)-benzene sulphonyl fluoride (AEBSF), 2 µg ml⁻¹ aprotinin, 7% glycerol, 2 mg ml⁻¹ phosphatidylcholine, Tris HCl (50 mM pH 7.4), [¹²⁵I]TRP (ref. 11; 120,000 c.p.m. per assay) with or without 300 nM unlabelled TRP, for 60 min at 4 °C.

Western blotting of immunoprecipitated extracts. Extracts from 1321N1 cells or male rat anterior hemipituitaries (preincubated with 10 µM AT_{II} for 10 min at 37 °C) were immunoprecipitated using polyclonal anti-ARF1/3 antibodies at 20 µl ml⁻¹ (peptide at 8 µg ml⁻¹), polyclonal anti-RhoA IgG at 5 µg ml⁻¹ (peptide at 20 µg ml⁻¹), rabbit anti-M₃ receptor serum⁶ (antigen M₃ receptor_{561–578}) (or non-immune rabbit IgG) at 2.5 µg ml⁻¹ and polyclonal anti-AT_I receptor IgG at 5 µg ml⁻¹ (antigen AT_I receptor_{15–24}; 20 µg ml⁻¹; Santa Cruz). Proteins collected by protein G beads were separated by SDS-PAGE and blots were incubated with monoclonal anti-ARF IgG (clone 26, 1:200 dilution; Transduction Laboratories), or anti-Rho IgG (clone 26C4, 1:200 dilution; Santa Cruz). Detection was by horseradish-peroxidase-conjugated secondary antibody and enhanced chemiluminescence.

Modulation of agonist affinity at M₃ receptors by GTP analogues. 1321N1 cells were incubated with or without carbachol (20 µM) and/or BFA (100 µM) for 10 min at 37 °C. Carbachol displacement (10–3,000 µM) of membrane [³H]NMS binding was measured in 10 mM MgCl₂, 100 mM NaCl, 20 mM Hepes pH 7.5, 1 nM [³H]NMS, with or without GTPγS, GPPCH₂P or F⁻, for 60 min at 37 °C.

[¹²⁵I]buserelin-binding studies in transfected COS 7 cells. Receptor constructs in pcDNA1 (refs 21, 24) were transfected using DEAE dextran (20 µg DNA per 4 × 10⁶ cells) and were assayed 72 h later. Membrane binding was assayed as described³⁰. For cell-surface binding, cells in 12-well plates were incubated with ligand, with or without 3 µM GnRH, at 4 °C or 37 °C as appropriate³⁰. Internalization was measured at 37 °C for 0–40 min after pre-equilibration with ligand for 90 min at 4 °C. Surface-bound ligand was dissociated with cold 0.2M acetic acid in 0.5M NaCl. Ligand dissociation in permeabilized cells was measured by prelabelling at 4 °C, and then successive incubations with medium containing 22 µM digitonin and GTP analogues at 37 °C. Initial dissociation occurred identically from wild-type and Asn 318 mutant GnRH receptors, reflecting a temperature-dependent reduction in GnRH binding. From 15–50 min, slower dissociation rates were reached and plots of ln B₀ against time revealed single linear components of half-lives: wild-type: 22 ± 3 min, and Asn 318 mutant: 26 ± 3 min in controls.

Solubilization and immunoprecipitation of wild-type and mutant GnRH receptors labelled by [¹²⁵I]buserelin. After preincubating cells with or without 100 nM GnRH for 15 min, membranes were solubilized in 5 mM CHAPS and 1.5M NaCl (ref. 30). Extracts were adjusted to 0.5M NaCl and

precleared before incubating (at 4 °C for 18 h) with the polyclonal antibodies anti-ARF1/3 (10 µl ml⁻¹) and anti-RhoA (1 µg ml⁻¹), described above, or with mouse monoclonal anti-Ras IgG (antigen p21^{ras}; Transduction Laboratories) or mouse monoclonal anti ERK1/2 IgG (antigen ERK1_{325–345}; Zymed Laboratories) as controls at 1 µg ml⁻¹ (both of which are reported to recognize native conformations of their targets). Blocking peptides for the polyclonal reagents were used at 2 µg ml⁻¹. Immune complexes were collected using protein G-Sepharose, and [¹²⁵I]buserelin binding was measured in polyethylene glycol precipitates³⁰.

Received 12 November; accepted 19 December 1997.

1. Brown, H. A., Gutowski, S., Moomaw, C. R., Slaughter, C. & Sternweis, P. C. ADP-ribosylation factor, a small GTP-dependent regulatory protein, stimulates phospholipase D activity. *Cell* **75**, 1137–1144 (1993).
2. Cockcroft, S. *et al.* Phospholipase D: a downstream effector of ARF in granulocytes. *Science* **263**, 523–526 (1994).
3. Exton, J. H. Phosphatidylcholine breakdown and signal transduction. *Biochim. Biophys. Acta* **1212**, 26–42 (1994).
4. Nieto, M., Kennedy, E., Goldstein, D. & Brown, J. H. Rapid heterologous desensitization of muscarinic and thrombin receptor-mediated phospholipase D activation. *Mol. Pharmacol.* **46**, 406–413 (1994).
5. Donaldson, J. G., Finazzi, D. & Klausner, R. D. Brefeldin A inhibits Golgi membrane-catalysed exchange of guanine nucleotide onto ARF protein. *Nature* **360**, 350–354 (1992).
6. Wall, S. J., Yasuda, R. P., Li, M. & Wolfe, B. B. Development of an antiserum against M₃ muscarinic receptors—distribution of M₃ receptors in rat tissues and clonal cell lines. *Mol. Pharmacol.* **40**, 783–789 (1991).
7. Rümenapp, U., Geiszt, M., Wahn, F., Schmidt, M. & Jakobs, K. H. Evidence for ADP-ribosylation factor-mediated activation of phospholipase D by m3 muscarinic acetylcholine receptor. *Eur. J. Biochem.* **234**, 240–244 (1995).
8. Zhang, S. J. *et al.* Rho-family GTPases regulate p38 mitogen-activated protein-kinase through the downstream mediator PAK1. *J. Biol. Chem.* **270**, 23934–23936 (1995).
9. Martin, A. *et al.* Activation of phospholipase D and phosphatidylinositol 4-phosphate 5-kinase in HL60 membranes is mediated by endogenous Arf but not Rho. *J. Biol. Chem.* **271**, 17397–17403 (1996).
10. Laudanna, C., Campbell, J. J. & Butcher, E. C. Role of Rho in chemoattractant-activated leukocyte adhesion through integrins. *Science* **271**, 981–983 (1996).
11. Feng, D. M., Veber, D. F., Connolly, T. M., Condra, C., Tang, M. J. & Nutt, R. F. Development of a potent thrombin receptor ligand. *J. Med. Chem.* **38**, 4125–4130 (1995).
12. Cavanagh, M. M. *et al.* Intracellular distribution of Arf proteins in mammalian cells. *J. Biol. Chem.* **271**, 21767–21774 (1996).
13. Bigay, J., Deterre, P., Pfister, C. & Chabre, M. Fluoride complexes of aluminium or beryllium act on G-proteins as reversibly bound analogues of the γ-phosphate of GTP. *EMBO J.* **6**, 2907–2913 (1987).
14. Geny, B. & Cockcroft, S. Synergistic activation of phospholipase D by protein kinase C- and G-protein-mediated pathways in streptolysin O-permeabilized HL60 cells. *Biochem. J.* **284**, 531–538 (1992).
15. Stutchfield, J. & Cockcroft, S. Correlation between secretion and phospholipase D activation in differentiated HL60 cells. *Biochem. J.* **293**, 649–655 (1993).
16. Houle, M. G., Kahn, R. A., Naccache, P. H. & Bourgoin, S. ADP-ribosylation factor translocation correlates with potentiation of GTPγS-stimulated phospholipase D activity in membrane fractions of HL60 cells. *J. Biol. Chem.* **270**, 22795–22800 (1995).
17. Malcolm, K. C., Elliott, C. M. & Exton, J. H. Evidence for Rho-mediated agonist stimulation of phospholipase D in Rat1 fibroblasts. *J. Biol. Chem.* **271**, 13135–13139 (1996).
18. Kahn, R. A. Fluoride is not an activator of the smaller (20–25 kDa) GTP-binding proteins. *J. Biol. Chem.* **266**, 15595–15597 (1991).
19. Gill, D. M. & Coburn, J. ADP-ribosylation by cholera toxin fragment A: functional analysis of a cellular system that stimulates the enzymic activity of cholera toxin fragment A. *Biochemistry* **26**, 6364–6371 (1987).
20. Hoffenberg, S. *et al.* Specific and effective interaction of a guanine nucleotide analogue with small G proteins. *Mol. Pharmacol.* **49**, 156–164 (1996).
21. Zhou, W. *et al.* A reciprocal mutation supports helix 2 and helix 7 proximity in the gonadotropin-releasing hormone receptor. *Mol. Pharmacol.* **45**, 165–170 (1994).
22. Zheng, L., Stojilkovic, S. S., Hundyady, L., Krstanovic, L. Z. & Catt, K. J. Sequential activation of phospholipase C and -D in agonist-stimulated gonadotrophs. *Endocrinology* **134**, 1446–1454 (1994).
23. Whatmore, J., Morgan, C. P., Cunningham, E., Collison, K. S. & Willison, K. R. ADP-ribosylation factor 1-regulated phospholipase D activity is localized at the plasma membrane and intracellular organelles in HL60 cells. *Biochem. J.* **320**, 785–794 (1996).
24. Sealfon, S. C. *et al.* Related contribution of specific helix 2 and 7 residues to conformational activation of the serotonin 5-HT_{2A} receptor. *J. Biol. Chem.* **270**, 16683–16688 (1995).
25. Chardin, P. *et al.* A human exchange factor for ARF contains Sec7-homology and pleckstrin-homology domains. *Nature* **384**, 481–484 (1996).
26. Morinaga, N., Tsai, S.-C., Moss, J. & Vaughan, M. Isolation of a brefeldin A-inhibited guanine nucleotide-exchange protein for ADP-ribosylation factor (ARF) 1 and ARF3 that contains a Sec7-like domain. *Proc. Natl Acad. Sci. USA* **93**, 12856–12860 (1996).
27. Randazzo, P. A. & Kahn, R. A. GTP hydrolysis by ADP-ribosylation factor is dependent on both an ADP-ribosylation factor GTPase-activating protein and acid phospholipids. *J. Biol. Chem.* **269**, 10758–10763 (1994).
28. Kwak, J. Y., Lopez, I., Uhlinger, D. J., Ryu, S. H. & Lambeth, J. D. RhoA and a cytosolic 50 kDa factor reconstitute GTPγS-dependent phospholipase D activity in human neutrophil subcellular fractions. *J. Biol. Chem.* **270**, 27093–27098 (1995).
29. Mitchell, R., Wolbers, W. B., Sim, P. & Fennell, M. The regulation of phospholipase C (PLC) and phospholipase D (PLD) by G protein receptor-activated tyrosine kinases in αT3-1 cells. *Biochem. Soc. Trans.* **23**, 208S (1995).
30. Ogier, S.-A., Mitchell, R. & Fink, G. Solubilization of a large molecular weight form of the rat GnRH receptor. *J. Endocrinol.* **115**, 151–159 (1987).

Acknowledgements. R.M., E.L. and M.J. are members of the Membrane Biology Group, University of Edinburgh. This work was funded by the MRC and the NIH. M.F. was supported in part by Wellcome Research Laboratories. We thank M. Wakelam for the polyclonal ARF antibody; G. Bokoch for RhoA constructs; B. Wolfe for M₃ receptor antiserum; the Scottish Antibody Production Unit for secondary antibodies; R. Clegg, L. Garland, N. Birdsall, B. Dickey, H. Weinstein, B. Ebersole, S. Dracheva and T. Harmar for help and advice; J. Bennie and S. Carroll for ligand iodination; and M. Eastwood for secretarial assistance.

Correspondence and requests for materials should be addressed to R.M. (e-mail: rmitchell@srv1.bmu.mrc.ac.uk).

To market, to market...

This roundup of recently announced products includes isoelectric focusing plates, filter capsules and electrophoresis kits for separations, as well as lamps, sample wheels and instruments for spectroscopy.

FinePoint and Pipet Plus

From Rainin

Help to eliminate aerosol contamination with a one-two combination of products

FinePoint aerosol-resistant tips provide a barrier against contamination during procedures such as PCR, sequencing, handling of infectious or radioactive samples. In case of overfilling, the hydrophobic filter allows the sample to be recovered by simply dispensing and recovering a particulate-free sample. Each lot is individually tested for RNase, DNase, DNA and pyrogen contamination, as well as undergoing sterilization by radiation after packaging. Unlike bevelled tips, the FinePoint wall thickness decreases gradually along the length of the tip. This widens the touch-off angle with the receiving vessel, which is said to result in more consistent touch off. Also useful against contamination are the Pipet-Plus latch-mode pipettes, which help to reduce aerosol generation and splashing by using an aspiration rate controller to control the uptake speed and improve consistency.

Reader Enquiry No. 100

The eyes have it

Pioneer 2000

From Iles Optical

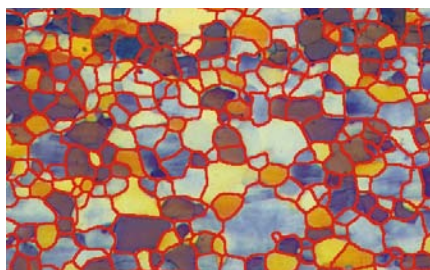
Eye and face protection that are said to incorporate advances in safety design

This range has been developed to offer a combination of style and economy. The twin-lens construction is said to give excellent forehead and cheek protection (UK kite-marked and CE approved to EN166). A wishbone side-arm design allows the adjustment of both the side-arm length and angle in a single movement for optimum fit to individual facial contours and maximum protection. The Pioneer 2000 has a comfortable raised, universal nosebridge for reduced lens abrasion. The new range is available in a choice of colours: black, amber, blue, metallic blue and metallic silver with lens materials and coatings to suit every application.

Reader Enquiry No. 101



Quite a spectacle: Iles Optical's new safety glasses.



Hard evidence: etched steel as seen by KS Grains.

KS Grains

From Imaging Associates

A new approach to grain sizing with the KS series of image analysis products from CZV

This is a dedicated software package designed for automatic grain sizing, which incorporates the advanced morphological techniques. Simple operating procedures and high-throughput are designed into KS Grains. In fact, the entire process, from image input to report, can sometimes be completed by simply clicking the 'do all' button. When difficult samples are encountered, decision support features can be introduced in the form of overlays, read from a library of known grain-boundary outlines. These overlays are superimposed on the image being evaluated, making this a truly automatic, interactive system. Sample data and evaluation output is stored in MS Access-compatible files and the system runs under either Windows95 or Window NT.

Reader Enquiry No. 102

Separation systems

TC SepraGels

From Integrated Separation Systems

Gels for the separation of low- to mid-range molecular weight proteins

These pre-cast Tricine-compatible gels differ from the traditional Laemmli formulation in their pH, buffer and cross-linker concentrations. They were developed for use with the Tris-Tricine-SDS buffer system to maximize resolution of low molecular weight proteins. In addition, these gels can be used in Tris-glycine-SDS running buffer to give a separation range that differs only slightly from the conventional Laemmli system. TC SepraGels are also compatible with Tris-glycine running buffer for native protein PAGE and TBE running buffer for nucleic acid PAGE. They are available as a 10–20 per cent gradient in both large and small formats. Custom orders are also available.

Reader Enquiry No. 103

IsoGel IEF

From Flowgen

Plates for protein separation and analysis using isoelectric focusing (IEF)

These plates are manufactured by FMC and have been designed for low cost and to save on labour. They are available in packs of six and are said to provide rapid and reproducible IEF of high molecular weight proteins (up to 2,000 K). Also, IsoGel plates are said to be well suited for the separation and analysis of antibodies. After IEF, the separated proteins can be transferred from the agarose gel to a nitrocellulose membrane, stained *in situ* or detected by antibodies within one hour. The agarose is said to be highly purified and to have no measurable EEO, owing to a custom manufacturing process that minimizes fixed anions and mobile cations. The agarose is cast on GelBond film to provide dimensional stability and ease of handling throughout processing.

Reader Enquiry No. 104

BioCap series

From Cuno

A series of disposable depth-filter capsules for process development

Now available in a suitcase pack containing 32 BioCap 30 filter capsules, this is a full range of Zeta Plus pharmaceutical-grade media that have different retention ratings. Process development scientists and engineers can quickly and easily optimize filter performance with this broad selection of Zeta Plus depth filter media, says Cuno. With 27 cm² of filter media, the capsules are well suited for scale-up evaluations as process volumes move from bench-scale to pilot-plant scale to manufacturing-scale operations. The performance of these capsules is said to be directly scalable to manufacturing operations with Cuno's 30- and 40-cm diameter filter cartridge systems. According to the manufacturer, Biocap filter capsules effectively remove cell debris, endotoxins, contaminating host-cell DNA, lipids, and other submicrometre contaminants from turbid, difficult-to-filter biological solutions. Primary applications for these filters include post-fermentation cell separation; the extension of life and protection of chromatographic columns; protection of downstream membrane filters; clarification of sera, reagents and buffers; vaccine broth clarification; and blood plasma fractionation.

Reader Enquiry No. 105

FACE electrophoresis kits

From Sigma

Complex carbohydrate and protein glycoconjugate characterization

Six kits for analysing complex carbohydrates and protein glycoconjugates, using the four-step FACE (fluorophore-assisted carbohydrate electrophoresis) technology, are now available from Sigma. Each kit includes all the necessary reagents and protocols to release, label, separate and image oligosaccharide or monosaccharide composition and sequence. Analyses can often be completed in just a single day, says Sigma. FACE should help in the analysis of post-translational modification and biological activity, as well as in the evaluation of expression systems and monitoring glycosylation. Among the kits available from Sigma are those for N-linked oligosaccharide profiling, N-linked oligosaccharide sequencing, O-linked oligosaccharide sequencing, monosaccharide composition analysis, glycosphingolipid oligosaccharide profiling, and glycosaminoglycan identification.

Reader Enquiry No. 106

Spectroscopy

HPK125 mercury lamp

From Cathodeon

A new high-pressure mercury lamp that adds to the company's range of line sources

This compact high-output lamp is stated to provide spectral irradiance of over 3,000 mW m⁻² nm⁻¹ at the major lines of 365 and 546 nm, as well as significant output at specific lines from 254 to 579 nm and a continuum output from 200 to 600 nm. The high-stability ultraviolet radiation, combined with an accurate position of the arc, makes the lamp suitable for a wide range of instrumentation, including fluorimeters, mercury analysers, medical and environmental equipment. In addition, the lamp may be used in a wide variety of photochemical processing applications.

Reader Enquiry No. 107

Sample wheel

From Nicolet

A computer-controlled, motorized FT-IR sampling accessory

Designed for unattended transmission analyses of multiple solid samples, the accessory has thirty 13-mm diameter sample positions. The 13-mm sample holders are suited to holding potassium bromide pellets, thin films, salt windows and 3M IR cards (enabling the automation of liquid analyses). The sample wheel should prove useful for busy quality assurance and quality control laboratories, where analysis of large numbers of solid samples is routine (analyses of up to 60 samples per hour are stated to be possible). The company also states that the wheel



Take a spin with the new FT-IR sample wheel.

may increase throughput by 20–50 per cent. A snap-in, snap-out mechanism allows wheels to be attached and removed easily. The entire accessory fits in the sample compartment with the cover closed, allowing the user to purge the sample compartment once for the entire run of the wheel. The sample wheel works in conjunction with Nicolet Omnic, Grams 386/32 operating software.

Reader Enquiry No. 108

Lightwave

From WPA

A new ultraviolet/visible spectrophotometer that uses diode-array technology

Featuring a space-saving profile and a low price, this instrument has fixed optics, no moving parts and a construction that is designed to ensure a long, maintenance-free life without any need for recalibration. Lightwave is said to provide instantaneous scanning for any routine application in the 200–825-nm range with simple menu-driven microprocessor controls and a large digital display of scans and data in a variety of formats. An integral cuvette tray provides a more convenient storage area, acting as a rack during analysis for improved throughput and reduced spillage.

Reader Enquiry No. 109

OSI 5 SP portable spectrometer

From ADC

A portable spectrometer with graphical display for analysing plant samples

Designed for the non-destructive analysis of plant and biochemical samples, the spectrometer is compact and uses microprocessor control. Introduced by plant science instrumentation manufacturer ADC, the spectrometer has a integral graphical display, allowing spectral plots to be displayed in real time. The instrument is said to provide fast acquisition and analysis of spectra at wavelengths from 220 to 400 nm, 220 to 730 nm or 310 to 900 nm, to an accuracy of 0.3 nm. A battery life of up to 8 h allows the spectrometer to replace cumbersome equipment that is confined to the laboratory with a system that can be used in the field. With an internal storage capacity of 256 KB RAM, a 3.5" disk drive and a digital output through an RS232 port, the instrument has

unlimited data storage. Moreover, operation is menu-driven through user friendly function keys. Monitoring such functions as intensity against wavelength, spectrum change through time and the comparison of samples to an input standard can be carried out using the spectrometer. Applications include fluorescence, chemiluminescence and bioluminescence, as well as analysis of plant canopy and ground cover, monitoring ambient light and pigment analysis.

Reader Enquiry No. 110

Lock box

Rad-Lock Box Beta

From Research Products International

Locking containers available for improved laboratory radiation safety

Containers of various styles and sizes have been designed to meet [US] state and federal requirements for security of radioactive material. Containers feature tamper-resistant hardware with an integral key lock and are made of 9.5-mm thick acrylic. They reduce emissions from high-energy β isotopes, such as those produced by ³²P.

Reader Enquiry No. 111

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

serotec

...better products, better service, better quality...

- Antibodies to Human CD's
- Antibodies to Mouse CD's
- Antibodies to Rat CD's
- Antibodies to Ovine CD's
- Antibodies to Bovine CD's
- Veterinary Research Antibodies
- Antibodies to Hormones
- Antibodies to Microbial Antigens
- Apoptosis Reagents
- Adhesion Molecule Products
- Cytokine & Growth Factor Reagents
- Synthetic Peptide Conjugates
- Flow Cytometry Reagents
- Hollow Fiber Bioreactors
- Cell Signalling Reagents
- Neurological Reagents
- Reagents for Immunology
- Reproductive Biology Reagents (Hormones)
- TCS Reagents
- Tissue Culture Products

FOR RESEARCH USE ONLY

Services: 22 Bedford Row, London EC1R 4DG, UK
Tel: +44 (0)1865 812700 Fax: +44 (0)1865 812701
Telex: 940101 Serotec Ltd Fax: +44 (0)1865 812701
Order: Medical Tel: +44 (0)1865 812701
www.serotec.co.uk

READER ENQUIRY NO. 111

Davids
Biotechnologie

Custom antibody production

Planning - production - after sales support

- chicken egg yolk antibodies
- fast - cost effective - very high yield -
- rabbit antibodies
- monoclonal antibodies
- development and production -

NEW - recombinant antigen service -

Ask for our catalog: Davids Biotechnologie
Karl Anselm Str.1A, D-93051 Regensburg
Germany, email: Davids.Bio@T-online.de
Phone: +49(0)171/4900308 Fax/Phone: +49(0)941/948228

READER ENQUIRY NO. 112

Action at the edges in cancer research

Hopes that therapeutic advances are just over the horizon are bringing increased funding to cancer research. Scientists joining from a variety of fields should find an environment very different from that of the past decade.

Potter Wickware

In cancer research, the scientist can productively study almost any area of basic biology, whether it be signalling, apoptosis, motility, membrane synthesis or structural biology, says Daniel Louvard, research director of the Institut Curie in Paris, and an investigator of membrane interfaces. Researchers have begun to discern simple unifying principles underlying the intricate complexity of the course of the disease in individuals.

The atmosphere in cancer research today owes much of its excitement to the growing understanding that cancer mechanisms are both diverse and unifying, and that the apparent contradiction is resolvable, says Louvard. As we now realize, it is some finite and knowable number of misbehaving biological processes combinatorially relating to one another in a malignant clone that produces the individual disease profile.

"We're developing a new way of thinking about biology that will lead to real advances. This is one of the compelling attractions of cancer research in the period we are entering," says Louvard. Scientific understanding combined with remarkable technological advances brings us closer to delivering better diagnosis and treatment. Innovations such as the DNA gene chip array, a technology being developed by Affymetrix of Palo Alto, California, and others, will allow the construction of a complete catalogue of a patient's genetic activity. Treatment can be tailored from that, based on profiles of which genes are behaving normally and which are aberrant.

New ways of thinking

Thea Tlsty, who studies genetic instability in breast cancer cell lines at the University of California in San Francisco, shares the excitement about the new avenues of investigation now opening up. One example is the use of *trans*-retinol as an antineoplastic agent. "Instead of using poisons to kill the cell and straining to mitigate side-effects in healthy cells, this approach causes cancer cells to, in effect, differentiate themselves out of existence," she explains. It has become important in clinical use, and is just as significant as an intellectual concept, in showing a new way to think about cancer.

To explore and exploit new ideas, Tlsty says the cancer researcher ought to have competence in imaging, bioinformatics and computing, as well as in the biological and chemical basics. Yet it is a mistake for a scientist to think too closely about techniques, because in research the relationships

between different regions of enquiry change rapidly. "These relationships are like droplets trying to coalesce. All the action is at the edges. Suddenly the two come together and there is no more interface — or rather there is a new interface that pops up somewhere else." For the student contemplating a career in cancer research, or any area of biology, the point is to develop the ability to think analytically, to formulate the query, and to plan an experimental strategy that will work. These skills never lose their relevance, no matter what the experimental context.

But what about the Malthusian crisis of expanding numbers of PhDs chasing limited numbers of faculty positions and grants? "Only by taking a narrow view of a scientific education do you run up against a limit in cancer research," says Tlsty. "The number of faculty slots in universities is limited. But there are many other areas: government, consulting on environment and hazards, research and development in industry, writing, tutoring venture capitalists, patent law. And the cancer research world is changing so rapidly that the educational and interpretive roles of the scientist are becoming especially important."

Tomorrow's challenge

Louvard also sees opportunity as he gazes into the future of cancer research, but cautions that progress will depend on an intellectual reorganization. "We have become adept at dissecting a problem into its component parts, but now we have learned to put the pieces back together. To do this we'll have to improve communication among the participating disciplines. This is the great challenge of tomorrow."

Cancer research awaits those who, like the

group which catalysed the revolution in molecular biology a generation ago, have the ability to connect apparently unrelated findings and see the world in some fundamentally different way. "Perhaps it will be the physicists again," Louvard speculates. □

Potter Wickware is a science writer in Oakland, California, USA.

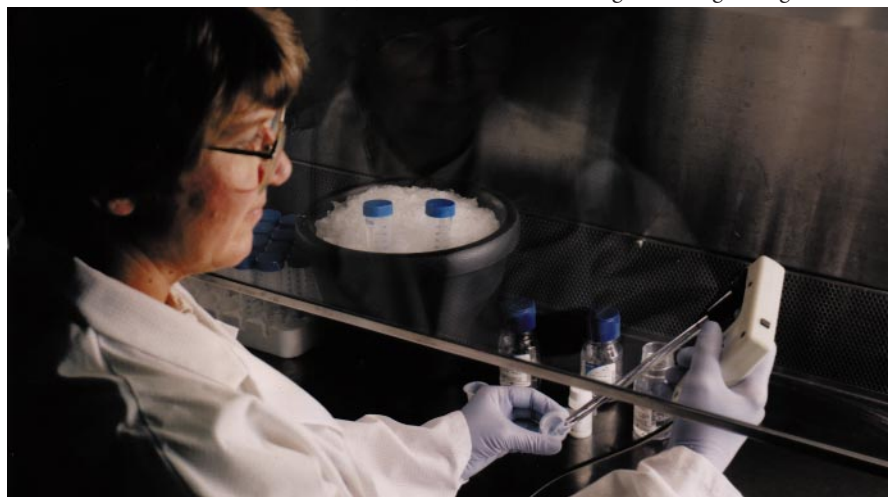
e-mail: wick@netcom.com

Seeking the bigger picture in the puzzle

Brendan Horton

Last December, Human Genome Sciences (HGS) of Rockville, Maryland, announced that it had applied to start phase I clinical testing of myeloid progenitor inhibitory factor-1. This compound, according to HGS, may allow oncologists to treat cancer patients with far more potent doses of chemotherapy. Screened from 300 full-length genes, this human protein has been shown in preclinical trials to "shield haematopoietic progenitor cells in the bone marrow from the effects of a number of chemotherapeutic agents used to treat all major cancers". HGS thinks the factor works by inhibiting proliferation and differentiation of these cells. Company chairman William Haseltine believes this to be the first genomics-derived therapeutic candidate to enter clinical testing. It will not be the last.

Not far from HGS, on the campus of the National Institutes of Health, the National Cancer Institute (NCI) has created a programme to advance the use of genomics and other technologies in diagnosing and treat-



Preparing a cDNA library: these will be a storehouse of genetic information for cancer researchers.

ing cancer. When Richard Klausner, now in his third year as NCI's director, joined the institute he took a broad look at its research funding and asked what could be done differently. According to his assistant, Robert Strausberg, five areas had little or no investment, so were all 'rescued' via the institute's bypass budget. These were cancer genetics, developmental diagnostics, detection technology, preclinical model development, and interfacing basic and clinical research (so-called translational research).

Genome anatomy project

Work groups were set up by Klausner to look at new ways to use cancer funding. From the developmental diagnostics group came the idea for a cancer genome anatomy project (CGAP). According to Strausberg, the group was dissatisfied that, despite much progress in learning about genes over the past ten years, the bigger picture was still missing from the cancer puzzle. They saw "an opportunity to take advantage of advancing technologies and genome resources in defining the complete molecular anatomy of cancer cells and their precursors," says Strausberg. The idea was to identify the best fingerprints that distinguish one cancer from another, as well as to differentiate the stages of cancer progression, even for "tumours that look identical histologically, but are different at the molecular level and respond differently to treatment".

Last year, two steps were taken towards making CGAP a reality. Both used funds from NCI's bypass budget. (The projects have since been written into this year's fiscal budget.) The first step was the creation of the tumour gene index, designed to house all the genes that are expressed during tumour development, covering the range from normal to precancerous and cancerous cells. Initially, five tumour types are being included: breast, prostate, lung, colon and ovarian.

Of primary importance to NCI is to put this information into the public domain as soon as it is received and verified. The dissemination of information and other resources is performed in collaboration with the IMAGE consortium project of cDNA library sequencing to find the expressed genes.

According to David Krizman, a molecular biologist in NCI's pathology laboratory, this technology is not yet highly evolved. It started

Table 1 Sites of Internet interest for cancer research

Site name	URLs: http://
General	
NCI's Cancer Genome Anatomy Project	www.ncbi.nlm.nih.gov/ncicgap/
The National Cancer Institute	www.nci.nih.gov/
Human EST project	genome.wustl.edu/est/esthmpg.html
NCI's technology transfer fellowship programme	www.nci.nih.gov/ttran/ttf/ttf.htm
Lawrence Livermore National Lab/IMAGE Consortium	www-bio.llnl.gov/bbrp/image/image.html
Corporations	
Human Genome Sciences	www.hgsi.com/
Life Technologies, Inc.	www.lifetech.com/
Stratagene	www.stratagene.com
Geron	www.geron.com/
Arctur	www.arctur.com/
Academic	
Patrick Brown's lab, Stanford	cmgm.stanford.edu/pbrown/
Stanford's Genome Resources	genome-www.stanford.edu/

with basic, pre-existing technology. Over the first six months, sources of cDNA libraries have included the NCI laboratory, Life Technologies of Gaithersburg, Maryland and Stratagene of La Jolla, California, as well as Bento Soares' laboratory at the University of Iowa. The libraries are arrayed at Lawrence Livermore National Laboratory, followed by sequencing at Washington University. To date 170,000 sequences have been deposited in GenBank by the CGAP project. Mapping of expressed sequence tags is performed by Stanford Genome system using radiation hybrid panels.

Construction of cDNA libraries is the near-term goal for CGAP, says Krizman. The longer-term goal is to increase usefulness through longer, more efficient inserts and to develop cost-effective, high-throughput, full-length cloning.

The second important aspect of CGAP is to create mechanisms for development of technologies that achieve "comprehensive high-throughput molecular analysis of gene and protein expression, as well as [techniques for] mutation detection". An early example is the development of the laser-capture microdissection (LCM) technique by Lance A. Liotta and colleagues at NCI. This allows the dissection of pure cells from different stages of cancer development, and permits libraries to be made that reflect these stages (see Table 1 for CGAP Web page).

Strausberg says it is important to make this technology available to the entire community and, for this reason, the technology transfer was rapid. It was commercially

developed as the PixCell LCM system through a Collaborative Research and Development Agreement partnership with Arcturus Engineering (see Table 1).

Future studies and technology

In addition to developing a wide variety of technologies for investigating changes in genes and their expression, an initiative has been announced that will create an interface between the technology developers and the clinical researchers (see Table 2). "As the technologies are being developed, we want researchers to think about the kinds of questions that could be answered with them," says Strausberg. One of NCI's main objectives is to build interdisciplinary teams of technology developers, including molecular biologists, engineers, physicists and cancer researchers, so that a complete system for analysis will be created.

Researchers would like to be able to trace the contribution that various genes make to the development of cancer. One aspect is identifying polymorphisms of genes that will allow the inheritance of the genes to be followed, says Strausberg. The NCI is building this capacity into the CGAP system, including the ability to identify and catalogue single nucleotide polymorphisms. "This is challenging and we are evaluating the technologies to do this efficiently," says Strausberg.

The institute is also beginning a mouse CGAP project as part of a programme to evaluate and standardize model organisms, as well as to look at analogous tumours between mice and humans for comparative analysis. Other technologies on the horizon, says Krizman, are those that allow investigation of mutation analysis, such as finding single-point mutations within coding sequences.

"There is an enormous opportunity, which we started catalysing last year with the NCI funds," says Strausberg. "You would like to gain more than just a few snapshots of the cancer, and to understand this as a dynamic process, to understand the events that happen in the transition from a normal cell to a cancer cell. We need to be able to look at

Table 2 Recent funding announcements from the NCI, the National Human Genome Research Institute (NHGRI) and the National Institute for Standards and Technology

1) High-throughput technologies to detect alterations in tumours	http://www.nih.gov/grants/guide/pa-files/PA-97-021.html
2) Novel technologies for evaluation of molecular alterations in tissue	http://www.nih.gov/grants/guide/rfa-files/RFA-CA-97-011.html
3) The development of genomic-scale technologies, or implementation of pilot-scale or large-scale projects for the discovery and scoring of single nucleotide polymorphisms	http://www.nhgri.nih.gov/Grant_info/Funding/rfa-hg-98-001.html
4) The development of a network for large-scale sequencing of the human genome	http://www.nhgri.nih.gov/Grant_info/Funding/rfa-hg-98-002.html

single cells, not merely populations of cells." This requires further development of the technology. For CGAP, having the funds and the ability to bring in a wider community of technology developers to work on these problems should expedite this process. □

Brendan Horton is on the staff of Nature, based in Washington. e-mail: b.horton@naturedc.com

Visionaries seek UK national strategy

Alison Mitchell

The Imperial Cancer Research Fund (ICRF) has a vision. By 2020, it believes that considerable advances will have been made in the prevention, early detection and treatment of cancer. But it cannot achieve these goals alone and, last month, it supported calls for a national strategy on cancer research — a collaboration between the UK government, the research charities and industry.

At present, charity research institutions cannot freely apply to government research agencies for support. Paul Nurse, director-general of the ICRF, says: "Government funding needs to be more flexible, to ensure that resources are spent where they can be best used." The stumbling block seems to be that government grant-awarding bodies allocate only a small percentage of their budgets to cancer research precisely because the

Table 3 Sites of Internet interest for UK cancer research

Site name	URLs: http://
General	
ICRF	www.icrf.net.uk/public.html
CRC	www.crc.org.uk
Leukaemia Research Fund	dSPACE.dial.pipex.com/lrf/
Academic	
Gray Laboratories	www.graylab.ac.uk/
CRC Beatson Laboratories	www.vet.gla.ac.uk/beatson/
Institute of Cancer Research	www.icr.ac.uk/
Marie Curie Research Institute	mc11.mcri.ac.uk/mcrihome.html
Paterson Institute for Cancer Research	christie.man.ac.uk/picr.htm
Ludwig Institute (UCL)	www.ludwig.ucl.ac.uk/
Ludwig Institute (St Mary's)	www.sm.ic.ac.uk/ludwig/
Corporations	
Serotec	www.serotec.co.uk/

cancer charities exist. But links are being forged between the charities and industry, and such schemes may provide opportunities for students, postdocs and group leaders.

The ICRF is the largest UK cancer charity, spending £54 million (US\$89 million) a year on research, closely followed by the Cancer Research Campaign (CRC) which spends £50 million. The charities distribute their funds very differently. The ICRF runs its own laboratories, while the CRC mainly supports research within universities and medical schools (see box below). How much room do these strategies leave for collaboration?

Plenty, it seems. Because of the way it distributes its funds, the CRC is particularly open to joint research ventures. As well as funding large groups within centres such as

the Institute of Cancer Research in London, in 1989 the CRC was one of the partners that set up the Wellcome/CRC Institute in Cambridge. The Wellcome Trust does not fund cancer research, but the scientific focus at this institute is on basic research into the cellular processes that are relevant to development. The ICRF's clinical directorate also has strong research partnerships with the National Health Service trusts. Mike Probert, the charity's assistant director of research, clinical division, says most of the clinical studies "would not be possible without the flow of patients through NHS hospitals".

The charities are also increasingly looking for ways to develop their discoveries commercially. The ICRF's technology transfer arm, Imperial Cancer Research Technology (ICRT), bridges the gap between basic research and product development. Last October, in a joint venture with the pharmaceutical company Antisoma, ICRT announced a phase III clinical trial of Theragyn. This drug is based on a murine monoclonal antibody developed by ICRF scientist Joyce Taylor-Papadimitriou and colleagues. The antibody is chemically linked to a radioactive isotope, yttrium-90, which is targeted to cancer cells. Phase I/II trials showed it to be particularly effective in patients with ovarian cancer.

Marketing teamwork

The CRC's technology transfer company is Cancer Research Campaign Technology (CRCT) which announced last month that it was teaming up with the UK company Serotec to market immunological reagents developed by the CRC. Another venture is Oncotech, which aims to develop the commercial potential of cancer research. This is a consortium of CRCT, the Leukaemia Research Fund, the law firm Cameron McKenna and the patent firm Mewburn Ellis.

So, in many ways, a national strategy for cancer research may not be very far away. Indeed, the ICRF and CRC already belong to a joint venture with the Medical Research Council and the Department of Health. But cancer research encompasses many areas of

Research opportunities at the cancer charities

Research at the ICRF is broken down into basic science, clinical studies and translational research — the middle ground between the two. Most of the basic science is carried out at the laboratories in central London and at Clare Hall in Hertfordshire.

These sites employ around 80 graduate students and 180 postdocs. Students usually receive funding for four years from the charity, and must register externally for a PhD.

About 30 postdoctoral fellowships are available each year in basic research. For those bringing their own grants, the ICRF offers access to support services (including cell and media production, oligonucleotides, instruments and antibodies). Recruitment of group leaders has slowed after big campaigns in the last

two years, but positions are available.

Clinical research units are based in hospitals and medical schools around the United Kingdom, covering areas ranging from immunotherapy to breast oncology. A total of 33 graduate students is funded by the ICRF, plus 44 clinical fellows and 49 postdocs.

The CRC divides its research between basic and clinical studies in a roughly 45:55 split. It awards grants to groups at the Paterson Institute in Manchester, the Beatson Institute in Glasgow, the Institute for Cancer Research and associated CRC centres, the Gray Laboratory Cancer Research Trust in London, and groups at universities such as Birmingham and Dundee.

Seventy-six students are funded by the CRC. Training

follows guidelines set by the host institutions.

The CRC runs three fellowship schemes — Research Fellowships for Clinicians, Senior Clinical Fellowships and Senior Cancer Research Fellowships. Between them, these support 597 postdoctoral fellows. Scientists can also apply for project grants, which are reviewed by the grants committee four times a year. The Wellcome/CRC Institute operates junior and senior research groups, with the junior groups staying for a maximum of ten years.

Both charities are developing 'translational research', to bridge the gap between basic and clinical research. The ICRF has a training scheme in this area, which is now in its third year, with ten clinical or postdoctoral fellows. **A.M.**

basic science — from cell growth and death to DNA repair and replication — as well as clinical trials, epidemiology and medical statistics. The ICRF and CRC cover all these areas, but there are also many smaller, more specialist charities. These many institutions have one common aim, however — to turn scientific advances into effective ways to prevent or treat cancer. □

Alison Mitchell is on the staff of Nature, based in London. e-mail: a.mitchell@nature.com

Therapeutics: a glimpse of the future

Owen Goldring

Cancer therapeutics has, until recently, been out of fashion — being seen as an applied rather than an academic discipline. But that is changing, thanks to advances in the understanding of the molecular nature of cancers, coupled with the hot debate centred around telomerase and the initial unravelling of the mechanisms of angiogenesis. A radical new generation of anti-cancer drugs should ensue.

The focus is on discovering small-molecule mechanism-based inhibitors that selectively target or starve cancers. There is a belief that there will be a first generation of *ras* farnesylation inhibitors, receptor tyrosine kinase inhibitors, and possibly p53 drugs, in the next five years (*ras* farnesylation and EGF receptor tyrosine kinase inhibitors are already in phase I clinical trials). Almost up with those will be the first angiogenesis inhibitors. The predictions are that cell-cycle inhibitors will be the next, with apoptosis after that. Telomerase inhibitors may come within five or ten years.

But, given that cancer is basically a disease of an unstable genome, some scientists believe that cancer will develop resistance to any drug thrown at it. They feel that more research is needed now to look at resistance mechanisms.

Search for selective drugs

There will be no sudden development of a 'pan-cancer' drug — the opposite will probably be true, with selective molecular drugs of high therapeutic index being used against specific solid tumours. Neither will conventional anti-cancer drugs, such as methotrexate and cyclophosphamide, be replaced overnight. But every researcher would like to offer a treatment that is more selective and considerably less toxic.

Drug discovery programmes are using techniques such as random screening for finding chemical leads, or structural rational design, where structural biologists use NMR and crystallography to figure out the molecular interaction between a target and an

inhibitor (see *Anti-Cancer Drug Design* 12, 525–531; 1997). These methods require a new breed of cancer scientist.

Paul Workman, the new director of the CRC Centre for Cancer Therapeutics, at the Institute of Cancer Research's labs in Surrey in the United Kingdom, says that, before the advent of molecular oncology, there was a cultural, ideological and skill-capability misfit between cancer biologists and those making derivatives of methotrexate. These two schools did not speak the same language. Now, they are beginning to, he says.

So, what skills do you need if you want to get into cancer therapeutics today? Workman says the type of scientist that biotech companies and even some major pharmaceutical companies lack is what he calls the "cancer pharmacologist". "These are not 'general' pharmacologists who are brought into a cancer project, as occurs in some companies. They are people who live and breathe cancer pharmacology, understand cancer pathways and can apply their pharmacological skills to these new molecular opportunities. They will not necessarily have cloned genes, but I do expect them to be able to do PCR, western blots, northern, Southern and so on." They need to understand the language of molecular biology, and traditionally trained pharmacologists tend not to have these skills, Workman says.

Two types of people are needed: those who have done biological science degrees, molecular biology, biochemistry, and then learn pharmacology; and those trained as traditional pharmacologists, because they bring a familiarity with concepts such as dose-response relationships and kinetics, says Workman. "Most molecular biologists have no training in these critical aspects of pharmacology. You need both. And you also need people trained in pharmacokinetics."

Angiogenesis link to mutation?

Each tumour needs to develop its own blood supply, says Roy Bicknell, head of the Imperial Cancer Research Fund's angiogenesis lab in the Institute of Molecular Medicine at John Radcliffe Hospital, Oxford. "Small tumours remain dormant until they undergo an angiogenic switch, when they start making angiogenic factors, and then recruit in new blood vessels. The tumour then grows rapidly. There is evidence to suggest the angiogenic switch is linked to a mutation in p53," says Bicknell.

Bicknell apparently does not mind which areas of science his PhD students come from: "They must be bright and inquisitive but, as long as they are practically adept, we reckon we can fill in any gaps in theoretical knowledge." One of his students has a first-class degree from Cambridge in zoology, another is a veterinarian from Germany, and a third has a medical degree from Oxford University.

Another, somewhat controversial, area of

cancer therapeutics is telomerase inhibition. About 85% of tumour cells produce telomerase, which keeps telomeres stable, and the assumption has been that cancer cells need telomerase to maintain continuous cell division.

Geron, a company based in Menlo Park, California, is studying the genetic clock of cell ageing, telomeres and telomerase. Geron has a substantial programme of research on telomerase inhibition (see Table 1).

Bioengineering angle

One possible route for future work on telomerase and angiogenesis is offered by Martin Braddock, leader of the endothelial gene expression group in the vascular diseases unit at Glaxo Wellcome's Stevenage labs in southern England: "When we expose endothelial cells in culture to fluid shear stress (for example, mimicking what happens with arterial blood flow), we have evidence that functional telomerase activity is downregulated." He encourages people to think about coupling telomerase activity with biomechanical forces, and relating that to cellular processes in angiogenesis.

"I think that's the way angiogenesis is going to go in the future," he says, adding that it will be understood almost from a bioengineering angle rather than from a purely cell or molecular biological angle. "We're almost at the start of the era of tissue engineering."

Cancer therapeutics is an area where the opportunity for doing exciting and groundbreaking research is very much on the up. But one problem that researchers will face — at least in Europe — is the dearth of second postdoc positions to enable them to get greater experience and a step further up the ladder to landing that elusive permanent academic job.

Many high-flyers gravitate from Europe to the United States after a first postdoc, do more academic research and then sometimes spend time in the biotech industry before returning to Europe. They are often welcomed back with open arms, being seen as having demonstrated an entrepreneurial outlook as well as having enhanced their academic experience.

Few people automatically turn to industry straight after a PhD or postdoc, perceiving that in industry people have to change subjects often. But this may be a misconception. Glaxo's Braddock says: "In our own particular area of research at the moment we are productive, and I hope that will remain so. If we hadn't been productive over the last few years we wouldn't work on it. But if we were in a university lab and non-productive, and had no grants to write after three years, the same situation would apply." □

Owen Goldring is a UK-based freelance science/medical writer and industrial consultant. e-mail: ogoldring@edit1.demon.co.uk.